2
2f

UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

FACULTAD DE CIENCIAS

FALLA DE ORIGEN

TEMAS SELECTOS DE PROBABILIDAD

T E S I S

QUE PARA OBTENER EL TITULO DE:

A C T U A R I O

P R E S E N T A :

ILEANA GARCIA CONDE

México, D. F. 1991

I N D I C E

INTRODUCCION

En el mundo que nos rodea, a cada momento ocurren fenómenos que no se pueden predecir con certeza; en ellos, se puede decir que existe un factor aleatorio ó de incertidumbre. La teoría de probabilidad tiene por objeto el estudio de este tipo de fenómenos; teniendo aplicación en casi todas las disciplinas del conocimiento humano. Es por esto, que nace la inquietud personal de presentar un camino intuitivo para conocerla y comprenderla.

Día con día las aplicaciones de la Probabilidad son más numerosas y profundas, no limitándose a solo las areas técnicas ya que se encuentran aplicaciones en areas de muy diversa índole. Así, por ejemplo, las podemos encontrar en la Aerodinámica, en la que se aplica en el estudio de los fenómenos de turbulencia, ó en el Cálculo Actuarial donde la probabilidad se constituye como un insumo básico.

La Teoría de Probabilidad tuvo sus origenes a mediados del siglo XVII, con hombres como Blaise Pascal, Pierre Fermat y James Bernoulli, en un principio se enfocó unicamente a los juegos de azar , sin embargo a mediados del siglo XVIII Pierre Laplace y Karl Friedrich hicieron la observación de que la teoria desarrollada para los juegos de azar podía ser aplicada a otras disciplinas, por ejemplo, haciendo una analogía entre estudiar los resultados "águila ó sol" y "vida ó muerte". A mediados del siglo XIX y principios del siglo XX, el desarrollo de la teoría esta asociada principalmente a Tchebychev, Markov, Liapunov quienes aportaron importantes ideas a la Ley de los Grandes Números. Posteriormente en 1917, Bernstein inició el proceso de axiomatizar la probabilidad, sin embargo, Kolmogorov formalizó la axiomatización relacionando los resultados existentes con la teoría de conjuntos y la teoría de funciones.

El propósito del presente trabajo es el de mostrar algunos resultados como apoyo para material didáctico de los cursos de Probabilidad I y Probabilidad II. Es conveniente mencionar que no son notas formales de la Teoría de Probabilidad ya que los temas aquí presentados consideran ante todo un punto de vista intuitivo y su generalización, se muestra no con exceso de teoría.

Entre los requisitos del lector para la comprensión de la tesis se encuentran los conceptos básicos de la teoría del álgebra y del cálculo.

La tesis consta de siete temas, en el primer capítulo se trabaja con el concepto de evento y algunas definiciones de Probabilidad concluyendo con la definición axiomática; asimismo se presenta el concepto de Probabilidad Condicional e Independencia.

En el segundo capítulo se definen a las variables aleatorias, y a la función de distribución. Las variables aleatorias se utilizan para describir a los eventos, y la función de distribución se utiliza para determinar la probabilidad de los eventos definidos en términos de las variables aleatorias.

El tercer capítulo se dedica a los vectores aleatorios, cuyos componentes son dos ó más variables aleatorias. Este capítulo es la generalización del capítulo anterior.

En algunas ocasiones por la naturaleza de los fenómenos aleatorios ó de la información que se tiene sobre de ellos es necesario evaluar la función de distribución de una función de variable aleatoria. Así en el cuarto capítulo se estudia las condiciones y la forma de evaluar esta función de distribución.

El quinto capítulo se dedica a las características númericas de las variables y vectores aleatorias (Esperanza Matemática, Varianza, Covarianza, Coeficiente de Correlación), y a la Función Generatriz de Momentos, la cual proporciona información muy utilizada en la descripción y análisis de las funciones de distribución.

En el sexto capítulo se da la definición de Función Caracteristica. Conociendo a esta función se puede determinar tanto los momentos de las variables y vectores aleatorios como la función de distribución de variables aleatorias definidas en términos de otras en las que se conoce su distribución.

El séptimo capítulo trata la convergencia de probabilidades, la convergencia de variables aleatorias y la convergencia de funciones de distribución; asimismo se exponen las Leyes de los Grandes Números y el Teorema del Límite Central.

En el apéndice se incluye la Función Generatriz de Momentos y la Función Característica de las Funciones de Distribución enunciadas en el segundo capítulo.

Quisiera expresar mi agradecimiento a todas las personas que me apoyaron con sus valiosos consejos y experiencias en la realización de este documento y en especial al Act. José Roberto Bautista quien estuvo a cargo de su dirección.

ESPACIO DE PROBABILIDAD

Día con día se encuentran fenómenos cuya ocurrencia es cierta, o bien es imposible. Sin embargo, cuando los fenómenos pueden ó no ocurrir bajo un conjunto de condiciones, como el que llueva ó no, toman el carácter de aleatorios y se llaman fenómenos ó experimentos aleatorios.

La Teoría de Probabilidad se encarga de estudiar dentro de un marco matemático formal, a este tipo de fenómenos.

Ejemplo:

Sea ε = experimento ó fenómeno aleatorio

ε_1 = (lanzar un dado),

ε_2 = (escoger al azar una letra del abecedario),

ε_3 = (lanzar una moneda),

ε_4 = (el número de coches que pasan por un semáforo),

ε_5 = (el número de aviones que no llegan a tiempo según itinerario),

ε_6 = (los casamientos en el mes de abril en la ciudad de México),

ε_7 = (el número de personas que mueren diariamente entre las 10 y 11 hrs.).

El estudio de los fenómenos aleatorios esta estrechamente ligado al conjunto de posibles resultados elementales del mismo; por lo que resulta de interés conocer a este conjunto, que se denota como Ω. Considérese de los ejemplos anteriores ε_1 = (lanzar un dado), como se conoce, existen 6 posibles resultados por lo tanto $\Omega_1=(1,2,3,4,5,6)$; entonces el conjunto Ω de los demás fenómenos enunciados es el siguiente:

$\Omega_2 = \{A,B,C, , ...X,Y,Z\}$,
$\Omega_3 = \{\text{águila, sol}\}$,
$\Omega_4 = \mathbb{N} \cup \{0\}$,
$\Omega_5 = \mathbb{N} \cup \{0\}$,
$\Omega_6 = \mathbb{N} \cup \{0\}$,
$\Omega_7 = \mathbb{N} \cup \{0\}$.

Si el fenómeno aleatorio (c) se repite un gran número de veces bajo las mismas condiciones, se descubriran ciertas regularidades que llevan a establecer que existe cierta frecuencia implícita en la ocurrencia de los resultados.

Una vez que se ha identificado el conjunto de posibles resultados de los fenómenos aleatorios los cuales son mutuamente excluyentes -solo sucede un único resultado-, surgen otras interrogantes, es decir, si

c = (lanzar un dado) *experimento o fenómeno aleatorio*
$\Omega = \{1,2,3,4,5,6\}$ *conjunto de posibles resultados*

entonces,

¿Obtendré un seis?,
¿Obtendré un número par?,
¿Obtendré un número menor ó igual a cuatro?,

así, pueden surgir una infinidad de interrogantes; por lo que para adentrarnos en el estudio de los fenómenos aleatorios se necesita construir una familia **A** de Ω, que contenga a todas las posibles preguntas sobre el experimento. Más formalmente:

Sea **A** una familia de subconjuntos de Ω, tal que cumple con las siguientes propiedades:

i) si $A \in \mathbb{A} \Rightarrow A^c \in \mathbb{A}$

ii) $\Omega \in \mathbb{A}$

iii) $\varnothing \in \mathbb{A}$

iv) si $(A_i)_{i=1}^{n} \in \mathbb{A} \Rightarrow \bigcup_{i=1}^{n} A_i \in \mathbb{A}$

v) si $(A_i)_{i=1}^{n} \in \mathbb{A} \Rightarrow \bigcap_{i=1}^{n} A_i \in \mathbb{A}$.

Siendo A un subconjunto de Ω, las únicas interrogantes a las que el fenómeno aleatorio puede responder, son preguntas de la forma ¿$w \in A$?; ¿el dos está en el subconjunto de los resultados menores ó iguales a cuatro?.

El fenómeno aleatorio c está completamente detallado por Ω, por lo que las preguntas relacionadas con c también deben de estarlo con Ω. Así, las preguntas asociadas al fenómeno aleatorio se pueden definir como,

Definición 1.1.

Si $A \in \mathbb{A}$, entonces se llama al subconjunto A un *evento* de Ω.

Si se toma el experimento c_1, un evento A puede ser:
(tirar seis), (tirar un número par), ó
(tirar un número menor o igual a 4).

Definición 1.2.

A la familia $\mathbb{A}$ se le llama *Campo de eventos de* Ω o bién un *Algebra de eventos de* Ω y en el caso de que n ((de *iv*) y *v*) tienda a infinito entonces a $\mathbb{A}$ se le conoce como un *σ-Campo de Eventos de* Ω o bien un *σ-Algebra de Eventos de* Ω.

Hasta el momento se ha determinado que para el estudio de los fenómenos aleatorios la Teoría de Probabilidad define a:

Ω = conjunto arbitrario no vacío de posibles resultados y representa a lo que estudia la Teoría de Probabilidad.

A = una familia de subconjuntos de Ω, un σ campo de eventos; representa en donde se realizará el estudio de los fenómenos aleatorios.

Ahora bien, ¿Cómo estudia la Teoría de Probabilidad a los fenómenos aleatorios?.

ALTERNATIVAS DEL "COMO"

DEFINICION DE PROBABILIDAD

Existen diversas definiciones de probabilidad, con el propósito de analizarlas, se han englobado en los siguientes cuatro grupos:

1. Definiciones en las que se toma a la probabilidad como una medida cuantitativa del grado de certeza del observador.
2. Definiciones que reducen el concepto de probabilidad al hecho de ser equiprobable.
3. Definiciones que toman su punto de partida en la frecuencia relativa de la ocurrencia de un suceso en un gran número de repeticiones.
4. Definición Axiomática.

Las definiciones que se engloban dentro del primer grupo son totalmente subjetivas ya que dependen del observador, por ejemplo, si en el desplegado de una revista aparece que el 7 de Abril del año 2000 lloverá en la ciudad de México, el observador de la Cd.

México considerará que difícilmente llueve en su ciudad en el mes de abril y pensará que no es probable el suceso; sin embargo puede ser que otra persona tenga otro punto de vista, por ejemplo un fluviólogo que por sus estudios considere que el suceso es probable. Lo anterior nos lleva a que cada observador tiene una apreciación diferente sobre el suceso, y la probabilidad estará influenciada por su conocimiento acerca del fenómeno y por el punto de vista de cada uno.

Las definiciones que se clasifican dentro del segundo grupo llevan a la definición Clásica, es decir,

Definición 1.3.

Sea n el número de elementos de Ω, $n<\infty$, si para cualquier evento $w\in\Omega$, la $P(w)=\frac{1}{n}$ entonces se define que para cualquier $A\in\mathbb{A}$,

$$P(A)=\frac{\text{núm. de elementos en A}}{\text{núm. de elementos en }\Omega}.$$

Para ejemplificar a este grupo de definiciones supóngase que se tiene una moneda neutra y se desea conocer la probabilidad de obtener una águila. El razonamiento sería: como sólo hay dos formas en las que cae una moneda, águila ó sol, entonces la probabilidad de que caiga águila es de 1/2.

Proposición 1.1.

La definición Clásica cumple con las siguientes propiedades:

1. Para toda $A \subset \Omega$ $\quad P(A) \geq 0$.
2. Para el evento Ω $\quad P(\Omega) = 1$,
   ya que $P(\Omega) = \frac{n}{n} = 1$.

3. Si el evento A es expresado como la suma de eventos mutuamente excluyentes B y C con $A, B, C \in \mathcal{A}$ entonces

$$P(A) = P(B) + P(C).$$

Demostración:

Solo se realiza la demostración de la tercera propiedad, debido a que las otras se cumplen claramente. Sea

b el número de eventos favorables en B, y
c el número de eventos favorables en C

como B y C son mutuamente excluyentes quiere decir que los eventos favorables para B no son eventos favorables para C pero hay b + c eventos favorables para el evento B + C = A entonces

$$P(A) = \frac{b + c}{n} = \frac{b}{n} + \frac{c}{n} = P(B) + P(C). \quad \blacksquare$$

Se establece que la terna $(\Omega, \mathbb{A}, P_{\text{clásica}})$ es un marco matemático formal para el estudio de los fenómenos aleatorios, sin embargo conlleva algunas limitaciones:

a) Ω conjunto de posibles resultados debe ser finito ($m<\infty$),
b) los resultados deben ser equiprobables,

con esta definición no se podría determinar la probabilidad de que un hombre muera antes de los 50 años o bien cuando se tiene un dado cargado como determinar la probabilidad de que salga par.

Las definiciones del tercer grupo estan asociadas al concepto de frecuencia relativa. Esta definición surge como una necesidad de resolver las limitaciones de la definición clásica. Para introducirnos dentro de esta considérese los siguientes ejemplos:

I) Supóngase que se tiene una caja con 3-pelotas idénticas marcadas con el 1,2,3 y se desea determinar la probabilidad de obtener cada una de ellas.

Sea c = (Escoger una pelota al azar , anotar el número que tiene y regresarla a la caja), entonces

$\Omega = (1,2,3)$; el experimento se repite n veces.

Sea $N_n(K)$ = el número de veces que la K'ésima pelota es seleccionada, supóngase que el experimento se repite 20 veces y se obtiene lo siguiente:

1,1,3,2,1,2,2,3,2,3,3,2,1,2,3,3,1,3,2,2

$N_{20}(1) = 5$, $N_{20}(2) = 8$, $N_{20}(3) = 7$

la proporción de ocurrencia de los eventos (Frecuencia Relativa) es:

$$\frac{N_{20}(1)}{20} = 0.25 \qquad \frac{N_{20}(2)}{20} = 0.40 \qquad \frac{N_{20}(3)}{20} = 0.35$$

por lo tanto; una manera intuitiva de establecer la probabilidad de ocurrencia es tomando el cociente:

$$f_n(A) = \frac{n_A}{n}$$

en donde

n = núm. de repeticiones del experimento,

n_A = núm. de veces que ocurre el resultado A en las n repeticiones del experimento.

Este cociente satisface las siguientes propiedades:

i) $0 \le f_n(A) \le 1$ debido a que $0 \le n_A \le n$

ii) $f_n(\Omega) = 1$

iii) $f_n(A+B) = f_n(A) + f_n(B)$.

II) Considérese que se tiene los nacimientos realizados en el estado de Chiapas en determinados años ; como lo muestra la siguiente tabla. ¿Cuál es la probabilidad de que nazca una niña?.

Sea c = (nacimientos en el estado de Chiapas)

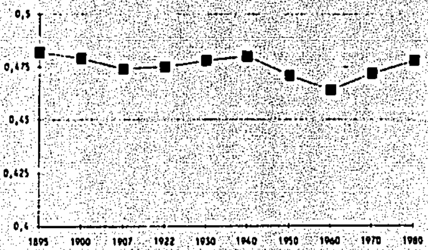
Ω = (hombres, mujeres)

A = (mujeres)

Nacimientos:

Años	Mujeres	Hombres	Total	$f_n(A)$
1895	2248	2415	4663	0.482
1900	5243	5699	10942	0.479
1907	6889	7618	14507	0.474
1922	5195	5732	10927	0.475
1930	10458	11385	21843	0.478
1940	12059	13020	25079	0.480
1950	16892	18962	35854	0.471
1960	22634	26111	48745	0.464
1970	28730	33102	60832	0.472
1980	37068	40131	77426	0.478

Gráfica



Se puede observar como $f_n(A)$ fluctúa alrededor de 0.475, por lo que se podría pensar que entre más grande es el número de repeticiones, la frecuencia relativa del experimento se acerca más a un núnero fijo $P(A)$ es decir:

$$\frac{n_A}{n} \underset{n\to\infty}{\rightarrow} P(A)$$

dado que $P(A)$ es una característica númerica del fenómeno es natural pensar que $P(A)$ es una medida de probabilidad del evento A.

Intuitivamente la frecuencia relativa se desvía cada vez menos de $P(A)$ conforme el número de experimentos se incrementa continuamente. Así, Von Mises[1] definió a la probabilidad (P) de un evento A como:

$$P(A) = \lim_{n\to\infty} \frac{\mu}{n}$$

en donde

μ = número de veces que el evento A ocurre en n repeticiones
n = número de repeticiones.

Proposición 1.2.

La definición Frecuentista cumple con las siguientes propiedades:

1. Para toda $A \subset \Omega$; $0 < P(A) < 1$
2. Para el evento cierto Ω ; $P(\Omega) = 1$
3. Si el evento A es expresado como la suma de eventos mutuamente excluyentes B y C con $A, B, C \in \mathcal{A}$ entonces

$$P(A) = P(B) + P(C).$$

[1] Gnedenko B.V.-The Theory of Probability, pág. 58.

Demostración:

1. $f_n(A) = \frac{n_A}{n} \approx P(A)$, como $0 \leq n_A \leq n$ entonces

$$0 \leq \frac{n_A}{n} \leq 1$$

por lo tanto

$$0 \leq P(A) \leq 1.$$

2. $P(\Omega) = \frac{n\Omega}{n}$, como $n\Omega = n$ entonces $\frac{n\Omega}{n} = 1$ y

$$\lim_{n\to\infty} \frac{n\Omega}{n} = P(\Omega) \text{ implica que}$$

$$1 = P(\Omega).$$

3. $P(A) = P(B) + P(C)$

$$\text{como } A = B + C \Rightarrow \frac{n_{B+C}}{n} = \frac{n_A}{n}$$

y como son mutuamente excluyentes

$$\frac{n_{B+C}}{n} = \frac{n_B}{n} + \frac{n_C}{n},$$

entonces

$$\frac{n_B}{n} + \frac{n_C}{n} = \frac{n_A}{n} \Rightarrow P(B) + P(C) = P(A).$$ ∎

La terna $(\Omega, \mathcal{A}, P_{relativa})$ es un marco matemático formal para el estudio de los fenómenos aleatorios, sin embargo conlleva algunas limitaciones:

a) Primero, ¿Cómo determinar que existe el límite de la sucesión $\frac{n_A}{n}$?. Segundo, si converge, ¿Cuantas veces habrá que repetir el experimento para estar cerca del límite?.
b) Depende de la intuición ya que hay que repetir los eventos bajo las mismas condiciones.

Definición Axiomática.

Dentro de la Teoría de las Matemáticas Modernas es común utilizar como axiomas a ciertas proposiciones que se presupone son ciertas por lo que no son demostradas dentro del desarrollo de la teoría. Por lo tanto, todas las demás proposiciones de la Teoría de Probabilidad se deducen lógicamente sobre la base de los axiomas adoptados.

En 1917, Serge N. Bernstein fue quién inició con el proceso de axiomatizar el concepto de probabilidad, pero fue preciso esperar hasta 1933, año en que A. N. Kolmogorov axiomatizó la teoría formalmente. Principalmente Kolmogorov relacionó la teoría de probabilidad con la teoría de conjuntos y la teoría de funciones. De este hecho la teoría empezó a tener un desarrollo inesperado en diferentes areas del conocimiento humano.

Definición 1.4.

La terna $(\Omega, \mathbb{A}, P)$ es un *espacio de probabilidad* si:

- Ω es un conjunto arbitario no vacío
- $\mathbb{A}$ es un σ campo de eventos ó elementos de Ω
- P es una función real $\mathbb{A} \to \mathbb{R}$.

Esta función P cumple con los siguientes axiomas:

1. $P(A) \geq 0 \quad \forall A \in \mathbb{A}$,
2. $P(\Omega) = 1$,
3. Si $(A_n)_{n=1}^{\infty} \in \mathbb{A}$ tal que $A_i \cap A_j = \emptyset \quad \forall i \neq j$ entonces

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n).$$

Se puede afirmar que esta definición de Probabilidad es una definición matemática, ya que establece que la probabilidad puede ser llamada función de probabilidad. Es importante destacar que no establece como es que P asigna valores al evento A. Este hecho, se analiza en el siguiente capítulo. Por ahora, con base en la definición anterior, se muestran ciertas propiedades que cumple P.

Proposición 1.3.

$$P(A^c)=1-P(A) \text{ para toda } A\in\mathcal{A}.$$

Demostración:

Se tiene que $\Omega=A\cup A^c$, como $A\cap A^c=\emptyset$ entonces $P(A\cup A^c)=P(A)+P(A^c)$, el primer miembro es $P(\Omega)$, por lo tanto $1=P(A)+P(A^c)$. ∎

El tercer axioma de la función P tiene como condición de que $A_i\cap A_j=\emptyset$ para toda $i\neq j$; sin embargo, puede ocurrir que se desee determinar la probabilidad de la unión sin que se cumpla esta condición. La siguiente proposición resuelve este problema,

Proposición 1.4.

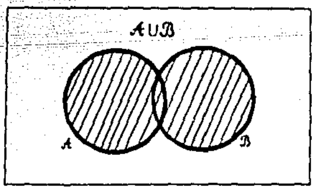
Para toda par de eventos (A,B), se cumple

$$P(A\cup B)=P(A)+P(B)-P(A\cap B).$$

Demostración:

Reescribiendo resulta $A\cup B=A\cup(B\cap A^c)$

Diagrama



entonces

$$P(A\cup B)=P(A)+P(B\cap A^c)$$

$$P(A\cup B)=P(A)+\underline{P(B\cap A^c)+P(A\cap B)}-P(A\cap B)$$

la parte subrayada es precisamente $P(B)$. ∎

El sistema de axiomas de Kolmogorov es consistente, dado que no se genera ninguna contradicción y existen objetos reales que satisfacen los axiomas. Por ejemplo, sea U un conjunto arbitrario conteniendo un número finito de elementos, es decir, $U=(a_1,a_2,...,a_n)$, y sea F el conjunto de todos los subconjuntos $(a_{i1},a_{i2},...,a_{is})$, $0\le i_1<i_2<...<i_s\le n$, $0\le s\le n$, de U. Se satisfacen todos los axiomas de Kolmogorov definiendo $P(a_i)=p_i\,(i=1,2,...,n)$, donde $p_1,p_2,...,p_n$ son números arbitrarios no negativos que cumplen con

$$p_1 + p_2 +...+ p_n = 1$$

y

$$P(a_{i1},a_{i2},...a_{is}) = p_{i1} + p_{i2}+...+ p_{is}.$$

Sin embargo, el sistema de axiomas es incompleto, debido a que existen proposiciones que no se pueden demostrar logicamente de los axiomas. Por ejemplo, dado un conjunto U se puede elegir en el conjunto F diferentes formas de evaluar la probabilidad. Esto es, en el ejemplo de lanzar un dado se puede establecer que

$$P(E_1) = P(E_2) = ... = P(E_6) = 1/6 \qquad (1)$$

ó bien

$$P(E_1) = P(E_3) = P(E_5) = 1/4 \qquad P(E_2) = P(E_4) = P(E_6) = 1/12, \qquad (2)$$

etc.

La incompletez de los axiomas en la Teoría de Probabilidad no quiere decir que la elección de los axiomas haya sido incorrecta sino que se deriva de la naturaleza misma de los eventos en donde para un mismo conjunto de resultados se tienen diferentes probabilidades asociadas. Por ejemplo, si se tiene un par de dados uno de los cuales es equiprobable y el otro no. En el primer caso la probabilidad esta dada por el sistema de ecuaciones (1), en el segundo caso por, "digamos", el sistema de ecuaciones (2).

Continuando con el progreso de la teoria, Kintchine dió una alternativa de axiomatización, cambiando solo la propiedad (3) de la axiomatización de Kolmogorov como sigue:

(3)' Para toda $(A_i)_{i=1}^{n} \in \mathcal{A}$, $A_i \cap A_j = \emptyset$ $i \neq j$ entonces

$$P(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i)$$

y para toda sucesión monótona decreciente en $\mathcal{A}$ $(A_1 \supset A_2 \supset \ldots \supset A_i \in \mathcal{A})$, tal que $\bigcap_{i=1}^{\infty} A_i = \emptyset$, entonces $P(A_i) \to 0$ cuando $i \to \infty$ A este resultado se le conoce como Axioma de Continuidad.

Sin embargo, Kolmogorov demostró que (3) y (3)' son equivalentes

Demostración:

Primero se analiza el hecho de que (3) $\Rightarrow$ (3)'

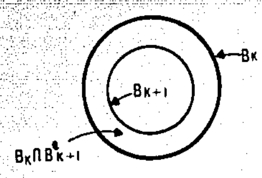
Sea una sucesión decreciente en $\mathcal{B}$ es decir

$B_1 \supset B_2 \supset B_3 \supset, \ldots\ldots$ tal que $\bigcap_{i=1}^{\infty} B_i = \emptyset$ entonces

P.D. $P(B_n) \to 0$ cuando $n \to \infty$.

$$B_i \cap B^c_{i+1} = B_i - B_{i+1}$$

como se aprecia en el siguiente diagrama



por consiguiente

$$B_n = \bigcup_{k=n}^{\infty} \left[B_k \cap B^c_{k+1} \right]$$

y además

$$(B_k \cap B^c_{k+1}) \cap (B_l \cap B^c_{l+1}) = \emptyset \qquad \forall\ k \neq l$$

entonces

$$P(B) = P \left[\bigcup_{l=1}^{\infty} \left[B_k \cap B^c_{k+1} \right] \right]$$

aplicando (3) de Kolmogorov resulta que

$$P(B_n) = \sum_{k=n}^{\infty} P(B_k \cap B^c_{k+1}) \tag{1}$$

(1) es la parte restante de la serie

$$P(B_1) = \sum_{k=1}^{\infty} P(B_k \cap B^c_{k+1})$$

es decir,

$$P(B_1) = \sum_{k=1}^{n-1} P(B_k \cap B^c_{k+1}) + P(B_n)$$

$$\lim_{n\to\infty} P(B_1) = \lim_{n\to\infty} \left[\sum_{k=1}^{n-1} P(B_k \cap B^c_{k+1}) \right] + \lim_{n\to\infty} P(B_n)$$

$$P(B_1) = \sum_{k=1}^{\infty} P(B_k \cap B^c_{k+1}) + \lim_{n\to\infty} P(B_n)$$

$$P(B_1) = P(B_1) + \lim_{n\to\infty} P(B_n)$$

por lo tanto

$$\lim_{n\to\infty} P(Bn) = 0.$$ ∎

Ahora, se verifica el regreso, es decir (3) ← (3)'.

Sea A_1, A_2, ...,A_n eventos tales que $A_i \cap A_j = \emptyset \ \forall \ i \neq j$ y $A = \cup A_i$.
Se define

$$B_n = \bigcup_{k=n}^{\infty} A_k = A_n \cup A_{n+1} \cup \ldots\ldots$$

es claro que

$$B_{n+1} \subset B_n$$

puesto que

$$B_{n+1} = A_{n+1} \cup A_{n+2} \cup \ldots\ldots$$

$$B_n = A_n \cup A_{n+1} \cup \ldots\ldots$$

si el evento B_n ocurrió, entonces para $i \geq n$, algún A_i ocurrió y ya que son excluyentes A_{i+1}, A_{i+2}, no pueden ocurrir. Esto es, los eventos B_{i+1}, B_{i+2},... son imposibles de ocurrir y consecuentemente el evento B_k es imposible de ocurrir por el Axioma de Continuidad $P(B_n) \to 0$ cuando $n \to \infty$, además

$$A = \bigcup_{i=1}^{n} A_i + B_{n+1}$$

por lo tanto

$$P(A) = P(A_1) + P(A_2) + P(A_3) + \ldots\ldots + P(A_n) + P(B_{n+1})$$

$$\lim_{n\to\infty} P(A) = \lim_{n\to\infty} \sum_{k=1}^{n} P(A_k) + \lim_{n\to\infty} P(B_{n+1})$$

entonces $P(A) = \sum_{k=1}^{\infty} P(A_k)$. ∎

PROBABILIDAD CONDICIONAL.

En algunas ocasiones se tiene que calcular la probabilidad de un evento A a condición de otro evento B cuya probabilidad de que ocurra es mayor que cero. A esta probabilidad se le conoce como probabilidad condicional y se denota como $P(A|B)$.

Definición 1.5.

Sea $(\Omega,\mathbb{A},P)$ un espacio de probabilidad, $A, B \in \mathbb{A}$, se dice que la probabilidad condicional de A dado B es:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \qquad \text{para toda } P(B) \neq 0.$$

Proposición 1.5.

Sea $(\Omega,\mathbb{A},P)$ un espacio de probabilidad, la terna $(\mathbb{B},\mathbb{A},P_B)$ donde $P_B(A) = P(A|B)$, es también un espacio de probabilidad.

Demostración:

$\mathbb{B}$ un conjunto arbitrario no vacio

$\mathbb{A}$ un α álgebra

P_B una función $P_B:\mathbb{A} \to \mathbb{R}$ tal que

i) Para toda $A \in \mathbb{A} \to P_B(A) \geq 0$,

ii) $P_B(\Omega) = 1$,

iii) Para toda $(A_i)_{i=1}^{\infty} \subset A$ con $A_i \cap A_j = \varnothing \quad i \neq j$

$$P_B\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P_B(A_i).$$

Solo se presenta la demostración de iii), debido a que las dos proposiciones anteriores se cumplen claramente.

Sea $(A_i)_{i=1}^{\infty} \subset \mathbb{A}$ tal que $A_i \cap A_j = \varnothing$ con $i \neq j$ entonces

$$P_B\left(\bigcup_{i=1}^{\infty} A_i\right) = P\left(\bigcup_{i=1}^{\infty} A_i \middle| B\right)$$

$$= \frac{P\left(\left(\bigcup_{i=1}^{\infty} A_i\right) \cap B\right)}{P(B)}$$

como $A_i \cap A_j = \emptyset$ si $i \neq j$, entonces $(A_i \cap B) \cap (A_j \cap B) = \emptyset$, si $i \neq j$; utilizando la hipótesis de $(\Omega, \mathcal{A}, P)$ es un espacio de probabilidad se tiene que:

$$\frac{P\left(\bigcup_{i=1}^{\infty}(A_i \cap B)\right)}{P(B)} = \frac{\sum_{i=1}^{\infty} P(A_i \cap B)}{P(B)} = \sum_{i=1}^{\infty} P(A_i | B) = \sum_{i=1}^{\infty} P_B(A_i). \quad \blacksquare$$

Ejemplo:

Supóngase que se tienen dos monedas idénticas, y se pretende:

a) Encontrar la probabilidad de que en ambas monedas caiga águila dado que en la primera ya cayó águila.

b) Encontrar la probabilidad de que en ambas monedas caiga águila dado que al menos una es águila.

En este caso Ω=(AA, AS, SA, SS), como las monedas son equiprobables, se definen los eventos

A=(la primera moneda sea águila),
B=(la segunda moneda sea águila)

$A = (AA, AS)$ $\qquad$ $B = (SA, AA)$ $\qquad$ $A \cap B = (AA)$

$P(A) = \frac{2}{4} = \frac{1}{2}$ $\qquad$ $P(B) = \frac{2}{4} = \frac{1}{2}$ $\qquad$ $P(A\ B) = \frac{1}{4}$

$A \cup B = (AA, AS)$ $\qquad$ $P(A \cup B) = \frac{3}{4}$

a) $P(A\cap B|A) = \dfrac{P(A\cap B\cap A)}{P(A)} = \dfrac{P(A\cap B)}{P(A)} = \dfrac{1/4}{1/2} = \dfrac{2}{4} = \dfrac{1}{2}$

b) $P(A\cap B|A\cup B) = \dfrac{P(A\cap B)}{P(A\cup B)} = \dfrac{1/4}{3/4} = \dfrac{4}{12} = \dfrac{1}{3}$.

INDEPENDENCIA.

Dos eventos A y B, son independientes cuando la realización de B no cambia la probabilidad de A. Es decir,

<u>Definición 1.6.</u>

Se dice que A y B son *independientes* si $P(A|B) = P(A)$.

Ejemplo:

Considérese el ejemplo anterior

$$P(A|B) = \frac{P(A\cap B)}{P(B)} = \frac{1/4}{1/2} = \frac{2}{4} = \frac{1}{2},$$

por otro lado $P(A) = \dfrac{1}{2}$ lo cual implica que A y B son independientes.

Proposición 1.6.

Se dice que A y B son independientes si y sólo si

$$P(A\cap B) = P(A)\ P(B).$$

Demostración:

$\Rightarrow$) A y B son independientes $\Rightarrow P(A|B) = P(A)$

$$\Rightarrow \frac{P(A\cap B)}{P(B)} = P(A) \quad \text{y como } P(B)\neq 0$$

$$\Rightarrow P(A\cap B) = P(A)\ P(B).$$

$\Leftarrow$) $P(A\cap B) = P(A)P(B)$ como $P(B)\neq 0$

$\dfrac{P(A\cap B)}{P(B)} = P(A) \Rightarrow P(A|B) = P(A)$, por lo tanto A y B son independientes. ∎

Ejemplo:

Supóngase que se tiran dos dados. Sea A el evento que denota que la suma de las caras sea impar, B el evento de obtener el número 1 en el primer dado, C es el evento de que la suma de las caras sea 7. Entonces ¿A y B son independientes?, ¿A y C son independientes?, ¿B y C son independientes?.

$$\Omega=\{(1,1),\ (1,2),\ (1,3),\ldots,(6,6)\}$$

$$P(A) = \frac{18}{36} = \frac{1}{2}, \qquad P(B) = \frac{6}{36} = \frac{1}{6}, \qquad P(C) = \frac{6}{36} = \frac{1}{6}$$

entonces

$$\frac{1}{12} = P(A\cap B)=P(A)P(B) = \frac{1}{2}\ \frac{1}{6} = \frac{1}{12},$$

$$\frac{1}{6} = P(A\cap C)\neq P(A)P(C) = \frac{1}{2}\ \frac{1}{6} = \frac{1}{12},$$

$$\frac{1}{36} = P(B\cap C)=P(B)P(C) = \frac{1}{6}\ \frac{1}{6} = \frac{1}{36},$$

por lo tanto A y B son independientes, A no es independiente de C y B y C son independientes.

Proposición 1.7.

Sean A y B eventos independientes entonces se cumple:

i) $P(A^c \cap B^c) = P(A^c)P(B^c)$

ii) $P(A^c | B) = P(A^c)$.

Demostración:

i) Se sabe que $A^c\cap B^c = (A\cup B)^c$ por consiguiente $P(A^c\cap B^c)=P((A\cup B)^c)$ y $P((A\cup B)^c)=1 - P(A\cup B)$ pero

$P(A\cup B)=P(A)+P(B)-P(A\cap B)$ como A, B son independientes entonces
$=P(A)+P(B)-P(A)P(B)$ en consecuencia,

$1-P(A\cup B)=1-P(A)-P(B)[1-P(A)]$ factorizando
$=[1-P(A)][1-P(B)]$
$=P(A^c)P(B^c)$.

ii) $P(A^c \cap B) = \dfrac{P(A^c\cap B)}{P(B)}$ pero $A^c\cap B=B-A\,B$ por consiguiente

$= \dfrac{P(B)}{P(B)} - \dfrac{P(A\cap B)}{P(B)}$, como A, B son eventos independientes entonces

$= 1 - \dfrac{P(A)\,P(B)}{P(B)} = 1 - P(A) = P(A^c)$. ■

Ahora bien, la anterior definición de independencia para dos eventos se puede generalizar de la siguiente manera:

Definición 1.7.

Sean $A_1, A_2, A_3, \ldots\ldots A_n$, n eventos, se dice que son independientes por parejas, si para toda $i \neq j$

$$P(A_i \cap A_j) = P(A_i)\ P(A_j).$$

Definición 1.8.

Sean $A_1, A_2, A_3, \ldots\ldots A_n$, n eventos, se dice que son mutuamente independientes, si para toda $i \neq j$

$$P(A_i \cap A_j) = P(A_i)\ P(A_j),$$

y para toda $i \neq j \neq k$

$$P(A_i \cap A_j \cap A_k) = P(A_i) P(A_j) P(A_k),$$

y así sucesivamente hasta

$$P(A_i) = \prod_{i=1}^{n} P(A_i).$$

De este modo, la probabilidad de n eventos independientes es igual al producto de las probabilidades de cada uno de los eventos. Esta definición no es redundante, es decir, que si son independientes por parejas no implica que para toda $i \neq j \neq k$ $P(A_i \cap A_j \cap A_k) = P(A_i) P(A_j) P(A_k)$.

Contraejemplo:

Sea $\Omega = (1,2,3,4)$ equiprobable

$A_1 = (1,2)$ $A_2 = (1,3)$ $A_3 = (1,4)$ entonces

$P(A_1) = P(A_2) = P(A_3) = \frac{2}{4} = \frac{1}{2}$, además,

$$\frac{1}{4} = P(A_1 \cap A_2) = P(A_1) P(A_2) = \frac{1}{2}\ \frac{1}{2}$$

$$\frac{1}{4} = P(A_1 \cap A_3) = P(A_1) P(A_3) = \frac{1}{2}\ \frac{1}{2}$$

$$\frac{1}{4} = P(A_2 \cap A_3) = P(A_2) P(A_3) = \frac{1}{2}\ \frac{1}{2},$$

por lo que A_1, A_2, A_3 son independientes por parejas sin embargo,

$$\frac{1}{4} = P(A_1 \cap A_2 \cap A_3) \neq P(A_1)P(A_2)P(A_3) = \frac{1}{2}\frac{1}{2}\frac{1}{2} = \frac{1}{8}.$$

De regreso tampoco se cumple,

$\Leftarrow$) Sea $\Omega = (1,2,3,4,5,6,7,8)$

$$A_1 = (1,2,3,4) \Rightarrow P(A_1) = \frac{4}{8} = \frac{1}{2}$$

$$A_2 = (4,5,6,7) \Rightarrow P(A_2) = \frac{4}{8} = \frac{1}{2}$$

$$A_3 = (4,2,5,8) \Rightarrow P(A_3) = \frac{4}{8} = \frac{1}{2}.$$

$$A_1 \cap A_2 \cap A_3 = (4) \Rightarrow P(A_1 \cap A_2 \cap A_3) = \frac{1}{8}, \text{ pero}$$

$$A_1 \cap A_2 = (4) \Rightarrow P(A_1 \cap A_2) = \frac{1}{8} \neq P(A_2)P(A_1) = \frac{1}{2}\frac{1}{2} = \frac{1}{4}$$

con lo que se demuestra que no se puede generalizar, que si se cumple para $i \neq j$ implique que se cumpla para n eventos.

Fórmula de Bayes.

El esquema general para aplicar la fórmula de Bayes en la solución de problemas prácticos es como sigue: Supóngase que un evento B puede ocurrir bajo un número n de diferentes condiciones A_i $(1 \le i \le n)$. Por una razón o por otra, las probabilidades de estas condiciones $P(A_i)$ son conocidas de antemano. Asímismo se sabe que las A_i asignan una probabilidad condicional $P(B|A_i)$ al evento B. Un experimento se realiza de tal forma que B ocurre, esto obligaría a una reexpresión de las probabilidades de las condiciones A_i dado que B ocurió; la fórmula de Bayes da una solución cuantitativa a este problema. Es decir, Proposición 1.8.

Sea $(A_i)_{i=1}^{n}$ una partición de Ω y sea B cualquier evento entonces $P(B) = P(B\cap\Omega) = P(B\cap[\bigcup_{i=1}^{n}(A_i)]) = P(\bigcup_{i=1}^{n}(A_i\cap B)) = \sum_{i=1}^{n} P(A_i\cap B)$ $= \sum_{i=1}^{n} P(B|A_i)P(A_i)$. A este resultado se le conoce como *Teorema de Probabilidad Total*. Ahora, si se quiere calcular $P(A_i|B)$ para alguna $i\epsilon(1,...,n)$ se tiene que:

$$P(A_i|B) = \frac{P(A_i\cap B)}{P(B)}$$
$$= \frac{P(A_i)P(B|A_i)}{P(B)}$$
$$= \frac{P(A_i)P(B|A_i)}{\sum_{i=1}^{n} P(A_i)P(B|A_i)},$$

a este resultado se le conoce como *Fórmula de Bayes*.

Para ejemplificarlo considérese lo siguiente. Un estudiante realiza un examén de opción múltiple, cada pregunta tiene 5 posibles respuestas de la cuales solo una de ellas es correcta. Si el estudiante conoce la respuesta correcta inmediatamente la selecciona, en caso contrario, selecciona al azar una de las 5 posibles opciones. Supóngase que el estudiante sabe la respuesta del 70% de las preguntas. Una pregunta interesante de conocer podría ser que si el estudiante escogió la respuesta correcta. ¿Cual es la probabilidad de que verdaderamente conozca la respuesta?.

$\Omega=(A_1, A_2)$ donde

A_1=(conocer la respuesta), y A_2=(no conocer la respuesta),

$P(A_1)=7/10$ y $P(A_2)=3/10$

Sea B=(obtener una respuesta correcta) entonces $P(B|A_1)=1$, $P(B|A_2)=1/5$ pero lo que se busca es $P(A_1|B)$, aplicando la fórmula de Bayes resulta

$$P(A_1|B)=\frac{P(A_1)P(B|A_1)}{\sum_{i=1}^{2} P(A_i)P(B|A_i)}=\frac{(7/10)(1)}{(7/10)(1)+(3/10)(1/5)}=\frac{35}{38}$$

por lo tanto la probabilidad de que un estudiante conozca la respuesta correcta dado que la contesto bien es de aproximadamente 92%.

En el presente capitulo se ha definido al espacio de probabilidad (Ω,A,P) como un marco matemático formal de los fenómenos aleatorios; sin embargo no se ha determinado como P asigna la "probabilidad" a los eventos. Esto se analiza en el siguiente capitulo.

VARIABLES ALEATORIAS Y FUNCIONES DE DISTRIBUCION

En el capítulo anterior se definió al espacio de probabilidad, representado por la terna $(\Omega, \mathcal{A}, P)$, como un marco matemático formal para el estudio de los fenómenos aleatorios. Sin embargo, no se estableció como es que la función de probabilidad asigna un valor al evento $A \in \mathcal{A}$. En este capítulo se definen a las variables aleatorias y a la función de distribución, las cuales se utilizan para determinar la probabilidad de los eventos definidos en términos de las variables aleatorias; lo anterior permite definir modelos que caracterizan a los fenómenos aleatorios.

VARIABLES ALEATORIAS.

Intuitivamente la *variable aleatoria* es una cantidad numérica que caracteriza al conjunto de posibles resultados del experimento aleatorio.

Supóngase que se lanza una moneda dos veces, por cada águila que cae se obtiene un peso, y por cada sol se pierde. ¿Cómo se calcularía la probabilidad de tener una ganancia de (-2),(0),(2) unidades?. El conjunto de posibles resultados se define como: Ω={(sol,sol), (águila,sol), (sol,águila), (águila,águila)}. Sea la variable X que representa la ganancia, entonces X puede tomar los valores {-2,0,2}, mismos que estan asociados al espacio muestral. La variable aleatoria X está asociando un número real a cada resultado del experimento entonces es natural pretender conocer la $P(X \leq -2)$, la $P(X \geq 2)$, o bien la $P(X=0)$. Más formalmente, la teoría debiera ser capaz de responder que para cualesquiera $x, y \in \mathbb{R}$:

$$¿P(w \in \Omega \; ; \; X(w) \leq x)?,$$
$$¿P(w \in \Omega \; ; \; X(w) > x)?,$$

$¿P(w \in \Omega \; ; \; X(w) = x)?$,

$¿P(w \in \Omega \; ; \; x < X(w) \leq y)?$ para $x<y$,

pero para dar respuesta a estas preguntas, se deberá pedir que se cumpla que para cualesquiera $x, y \in \mathbb{R}$

$$(w \in \Omega \; ; \; X(w) \leq x) \in \mathbb{A}$$
$$(w \in \Omega \; ; \; X(w) > x) \in \mathbb{A}$$
$$(w \in \Omega \; ; \; X(w) = x) \in \mathbb{A}$$
$$(w \in \Omega \; ; \; x < X(w) \leq y \;) \in \mathbb{A}.$$

Matemáticamente una variable aleatoria es un función medible definida sobre un espacio muestral. Esto se formaliza en la siguiente definición,

<u>Definición 2.1.</u>

Una *variable aleatoria* X en un espacio de probabilidad $(\Omega, \mathbb{A}, P)$ es una función de Ω en $\mathbb{R}$ tal que para toda $x \in \mathbb{R}$, $(w \in \Omega | X(w) \leq x) \in \mathbb{A}$.

El punto importante de esta definición es que si se piensa en término de los fenómenos aleatorios, Ω es el conjunto de posibles resultados y la variable aleatoria con dominio Ω hace que algún número real corresponda a cada resultado del fenómeno.

Considérese el ejemplo anterior, para verificar que X es variable aleatoria se debe checar que para toda $x \in \mathbb{R}$, $(w \in \Omega | X(w) \leq x) \in \mathbb{A}$ entonces,

$\mathbb{A}=\{$(sol,sol), (águila,sol), (sol,águila), (águila,águila)$\}$

$x<0$, $(w:X(w) \leq x)=($(sol,sol)$)$

$0 \leq x<1$, $(w:X(w) \leq x)=($(águila,sol), (sol,águila)$)$

$x \geq 1$, $(w:X(w) \leq x)=($(águila, águila)$)$

por lo tanto X es una variable aleatoria.

La variable aleatoria X se puede clasificar en discretas, continuas ó mixtas. En el presente trabajo solo se estudian a las dos primeras.

Definición 2.2.

i) Se dice que X es una *variable aleatoria discreta* si el rango de la función es un conjunto a lo más numerable de valores.

ii) Se dice que X es una *variable aleatoria continua* si existe una función f_X no negativa tal que $F_X = \int_{-\infty}^{X} f_X(u)du$ para cualquier valor de x.

Ejemplos:

Discreta. El ejemplo anterior, de lanzar dos dados,
sacar una pelota al azar de una urna,
comprar un billete de loteria.

Continua. La longitud del salto de una rana,
el tiempo que tarda una persona en contestar el teléfono,
seleccionar un punto al azar del intervalo (-5,5).

Como los resultados de los fenómenos aleatorios son inciertos, entonces los valores que toma la variable aleatoria también lo son; por lo tanto el conocer que valores toma la variable aleatoria no dice nada acerca de su ocurrencia, por lo que se introduce el concepto de Función de Distribución Acumulada, el cual permite asignar probabilidades a los valores que toma la variable aleatoria en forma acumulada.

FUNCION DE DISTRIBUCION.

Definición 2.3.

Sea $F:\mathbb{R} \to \mathbb{R}$ tal que cumple con las siguientes propiedades:

i) $F_X(x)$ es no decreciente, "es decir que para cualesquiera $x_1, x_2 \in \mathbb{R}$ tal que $x_1 \le x_2$ entonces $F_X(x_1) \le F_X(x_2)$"

ii) $F_X(\infty) = \lim_{x\to\infty} F_X(x) = 1$ y $F_X(-\infty) = \lim_{x\to-\infty} F_X(x) = 0$

iii) $F_X(x)$ es continua por algún lado (en particular se trabajará continua por la derecha),

entonces a F se le llama *Función de Distribución*.

Una función así definida, ayuda a responder a las preguntas asociadas al fenómeno aleatorio, es decir;

Proposición 2.1.

Sea $F : \mathbb{R} \to \mathbb{R}$ definida por

$$P(X \le x) = F_X(x) \qquad -\infty < x < \infty.$$

entonces, $F_X(x)$ así definida es la función de distribución acumulada de la variable aleatoria X en el punto x.

Demostración:

i) Sean $x_1 \le x_2$; P.D. $F_X(x_1) \le F_X(x_2)$.

Se define $A = \{w \in \Omega | x_1 < X(w) \le x_2\}$, entonces

$$P(A) = P\{w \in \Omega | x_1 < X(w) \le x_2\},$$

el punto a demostrar es que

$$P(w\in\Omega|x_1<X(w)\le x_2)=F_X(x_1)-F_X(x_2)$$

Sean $B=(w\in\Omega|X(w)\le x_2)$ y $C=(w\in\Omega|X(w)\le x_1)$ se tiene que $A\cup C=B$ y $A\cap C=\emptyset$ entonces $P(B)=P(A\cup C)=P(A)+P(C)$ pero $P(B)=F_X(x_2)$, $P(C)=F_X(x_1)$, por lo que $F_X(x_2)=P(x_1<X(w)\le x_2)+F_X(x_1)$ por lo tanto $F_X(x_2)-F_X(x_1) = P(x_1<X(w)\le x_2)$, lo cual implica que es no decreciente.

ii) $Fx(\infty) = 1$ y $Fx(-\infty) = 0$
$Fx(\infty) = P(w\in\Omega|X(w)\le\infty)=P(\Omega)=1$ abusando de la notación
$Fx(-\infty) = P(w\in\Omega|X(w)\le-\infty)=P(\emptyset)=0$ abusando de la notación.

iii) P.D. $F_X(x)$ es continua por la derecha.
Sea $x_n \to x$ una sucesión decreciente que converge a x por la derecha. Se define $A_n = (w \in \Omega | x < X(w) \le x_n)$, por el Axioma de Continuidad estudiado en el Capítulo I, $\lim_{n\to\infty} P(A_n)\to 0$, entonces $\lim_{n\to\infty} F_X(x_n) - F_X(x) \to 0$, por lo tanto $F_X(x_n) \to F_X(x)$. ∎

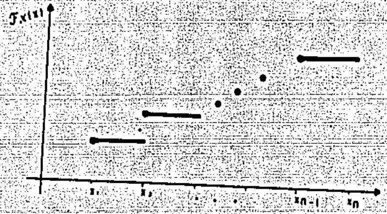
Ahora, el problema que se presenta es el de evaluar la función $F_X(x) = P(X\le x)$; primero se analiza el caso de las variables aleatorias discretas.

<u>Definición 2.4.</u>

Sea X una variable aleatoria discreta, es decir que toma los valores $x_1,x_2,x_3,\ldots$ entonces $F_X(x)$ es una función de distribución acumulada discreta si

$$F_X(x)=\Sigma_{x_i\le x} P(X = x_i).$$

La función de distribución $F_X(x)$ tiene en el punto $x = x_i$ el mismo valor que para toda x en el intervalo $x_i \le x < x_{i+1}$. Si x alcanza el valor x_{i+1} entonces la función de distribución aumenta, tal y como se observa en la siguiente gráfica



El valor de la función de distribución en el punto de discontinuidad es igual a su valor a la derecha, o sea, la función de distribución es continua a la derecha.

Definición 2.5.

Se dice que la función $f: \mathbb{R} \to \mathbb{R}$ definida por $f_X(x) = P(X=x)$ es la *función de densidad* de la variable aleatoria discreta X.

Proposición 2.2.

La función de densidad de la variable aleatoria discreta cumple con las siguientes propiedades:

i) $f_X(x) \ge 0$, para toda $x \in \mathbb{R}$
ii) $f_X(x) \ne 0$ en un número numerable de valores de $\mathbb{R}$
iii) $\Sigma f_X(x_i) = 1$.

Demostración:

i) $f_X(x)=P(X=x)\geq 0$, $x\in\mathbb{R}$. Como $(X=x)$ es un evento, por el primer axioma de probabilidad, $P(X=x)\geq 0$ para toda $x\in\mathbb{R}$.

ii) Dado que X, es una variable aleatoria discreta, entonces toma un número numerable de valores, por lo que se verifica que $f_X(x)=P(X=x)\neq 0$ en un número numerable de valores de $\mathbb{R}$.

iii) $\Omega = \bigcup_n (w:X(x)=x_n) = \bigcup_n (X=x_n)$ y $(X=x_j)\cap(X=x_i)=\emptyset$ para todo $i\neq j$ resulta que $1 = P[\Omega] = P[\bigcup_n (X=x)] = \sum P[X=x_n]$ por el tercer axioma de probabilidad. ∎

Por otro lado, se tiene que en el caso de que X sea una variable aleatoria continua, lo natural sería que

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\, dt$$

donde $f_X(t)$ es una función real; pero si esto es cierto, entonces por el teorema fundamental del cálculo se tendría que pedir que:

$$F'_X(x) = f_X(x)$$

y esto no necesariamente ocurre en las funciones continuas por la derecha; para aquellas funciones que si cumplan esta propiedad, se tiene la siguiente definición.

<u>Definición 2.6.</u>

Sea X una variable aleatoria continua, se dice que X es absolutamente continua si existe una función $f:\mathbb{R}\to\mathbb{R}$ tal que f es no negativa, es decir, $f_X(x)\geq 0$ para toda $x\in\mathbb{R}$,

$$\int_{-\infty}^{\infty} f_X(x)dx=1,$$

y para cualquier $x\in\mathbb{R}$

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\,dt$$

a la función $f_x(t)$ se le llama *función de densidad de probabilidad* de la variable aleatoria X.

PRINCIPALES MODELOS DE VARIABLES ALEATORIAS DISCRETAS Y CONTINUAS.

Se analizan los ejemplos más característicos, a fin de reafirmar las ideas expuestas.

<u>Variables Aleatorias Discretas.</u>

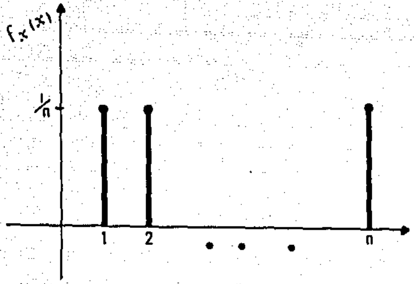
(*i*) Supóngase que se tiene una baraja de 52 cartas se quiere conocer ¿Cuál es la probabilidad de obtener un as si se toma una carta al azar?.

Ω=(52 cartas de la baraja),
La variable aleatoria es X=as, entonces
$P(X=\text{as})= 1/52$.

El experimento se describe con el siguiente módelo:

$$f_X(x) = \begin{cases} \frac{1}{n} & \text{si } x = 1,2,...n \\ 0 & \text{en otro caso} \end{cases}$$

entonces se dice que x variable aleatoria se distribuye uniformemente con parámetro *n*. Se denota como X-U(*n*) y su gráfica es la siguiente:



ii) Supóngase que se tiene un experimento aleatorio cuyo resultado esta clasificado en dos categorias éxito o fracaso, la variable aleatoria X toma los valores de 1 ó 0 con probabilidad *p* y $1-p=q$ respectivamente entonces

$$f_X(x) = \begin{cases} p^x(1-p)^{1-x} & x=0,1 \\ 0 & \text{en otro caso} \end{cases}$$

entonces X variable aleatoria tiene función de densidad *Bernoulli* con parámetro *p*. Se denota como $X\sim B(p)$ y su gráfica es como sigue:

Ejemplo:

Supóngase que se lanzan dos dados equiprobables; ¿Cuál es la probabilidad de que la suma sea 7?.Se asigna 1 si el resultado es el número 7 y cero en los demás casos. Entonces $p=6/36=1/6$ y $q=1-p=5/6$ entonces,

$$P(Exito) = f_X(1) = p^1(1-p)^0 = 1/6,$$
$$P(Fracaso) = f_X(0) = p^0(1-p)^1 = 5/6.$$

iii) De una sucesión de n ensayos independientes de Bernoulli, se quiere conocer el número total de éxitos, estos pueden ser 0,1,...n, el problema que se presenta es el de determinar la probabilidad correspondiente. El evento "de n ensayos resultan k éxitos y n-k fracasos" puede ocurrir del mismo número de maneras en que se distribuyen *k* éxitos en *n* lugares. Es decir, el evento contiene $\binom{n}{k}$ posibilidades, y por definición cada posibilidad tiene la probabilidad $p^k(1-p)^{n-k}$ ya que los ensayos son independientes, la función de densidad queda como sigue:

$$f_X(x) = \begin{cases} \binom{n}{x} p^x(1-p)^{n-x} & x=0,1,\ldots,n \\ 0 & \text{en otro caso} \end{cases}$$

entonces X variable aleatoria se dice que tiene función de densidad *Binomial* con parámetros (n,p). Se denota como $X\sim B(n,p)$.

Ejemplo:

Supóngase que de un lote de 10 focos 4 son defectuosos. Se toma un foco al azar, se anota su estatus y se regresa a la muestra, se realizan 20 repeticiones independientes; ¿Cuál es la probabilidad de obtener 5 focos defectuosos?.

Considérese que un foco defectuoso representa un éxito, entonces

$$p=P[\text{no defectuoso}] = \frac{10-4}{10} = 0.6,$$

$$q=P[\text{defectuoso}] = 1 - \frac{10-4}{10} = 0.4;$$

sea X la variable aleatoria que representa el número de focos defectuosos, el número de repeticiones es 20 entonces

$$f_X(5)= \binom{20}{5} p^5(1-p)^{20-5} = .001294,$$

lo que implica que la probabilidad de encontrar 5 focos defectuosos en 20 repeticiones es muy baja.

v) En los ensayos independientes de Bernoulli, (éxito y fracaso) en donde la n es comparativamente grande, la *p* pequeña y el producto ($\lambda=np$) de magnitud moderada, se puede usar una aproximación debida a Poisson. Esto es, sustituyendo resulta que,

$$p(x) = \binom{n}{x} p^x(1-p)^{n-x} = \frac{n(n-1)\ldots(n-x+1)}{x!}\left(\frac{\lambda}{n}\right)^x\left(1-\frac{\lambda}{n}\right)^{n-x}$$

$$\text{si } n\to\infty,\quad \frac{n(n-1)\ldots(n-x+1)}{n^x} \to 1 \text{ además } \left(1-\frac{\lambda}{n}\right)^{-x} \to 1$$

y por el desarrollo de la serie de Taylor

$$\left(1-\frac{\lambda}{n}\right)^n \to e^{-\lambda}$$

por lo tanto

$$p(x) \to \frac{e^{-\lambda}\lambda^x}{x!}.$$

Para poder aplicar esta aproximación hay que tener cuidado en que el valor de la *p* sea verdaderamente pequeño en comparación con n y el producto de magnitud moderada. En el ejemplo anterior de los focos la aproximación no es muy buena ya que λ=12 y p=0.6 entonces,

$$p(5) \to \frac{e^{-12}12^5}{5!} = 0.012740$$

que varia considerablemente del resultado obtenido.

Sin embargo, si la probabilidad de que sea un foco defectuoso es p=0.015, la probabilidad de que un lote de 100 focos no contenga ninguno defectuoso es $(1-0.015)^{100}$=0.2261. La aproximación de Poisson correspondiente es $e^{-1.5}$ = 0.22313, que es satisfactoria.

Debido a su importancia se formaliza como sigue:

X variable aleatoria tiene función de *densidad Poisson* con parámetro λ si

$$f_X(x) = \begin{cases} \dfrac{\lambda^x e^{-\lambda}}{x!} & x=0,1,2,... \\ 0 & \text{en otro caso} \end{cases}$$

con λ=np. Se denota como $X \sim P(\lambda)$.

Los fenómenos aleatorios para los cuales existe una clase de ordenamiento son candidatos de ser modelados por la función de distribución *Poisson*; es decir fenómenos en los que se observa la ocurrencia de ciertos sucesos durante el tiempo, el espacio, región o distancia y la λ es un promedio.

Ejemplo:

En cada tiro de arco, la probabilidad de dar en el blanco es 0.001. Encontrar la probabilidad de acertar con dos o más flechas si el total de ellas es de 5000.

En este caso $\lambda=np=(5000)(0.001)=5$, y para calcular la probabilidad deseada, se resta de 1 la probabilidad de acertar cero y una vez, es decir,

$$P(X\geq 2)=1-P(X=0)-P(X=1),$$

por la función de densidad de Poisson $P(X\geq 2) \cong 1-e^{-5}-5e^{-5} \approx 0.9596$, lo cual indica que es bastante probable de acertar con dos ó más flechas en el blanco.

vi) Considérese una sucesión de n ensayos independientes de Bernoulli. Se experimenta hasta que se obtiene el r-ésimo éxito, r es un número positivo; sea X variable aleatoria que represente el número de fracasos antes del r-ésimo éxito, entonces

$$P(X=x)=P\left\{(r-1) \text{ éxitos en los primeros } r+x+1 \text{ intentos}\right\}$$
$$* \; P\left\{(r+x) \text{ intentos resulte un éxito}\right\}.$$

Se sabe que

$$\binom{x+r-1}{r-1} p^{r-1}(1-p)^{x}\, p = \binom{r+x-1}{x} p^{r-1}\,(1-p)^{x} p,$$

por lo que resulta

$$f_X(x) = \begin{cases} \binom{r+x-1}{x} p^{r}(1-p)^{x} & x=0,1,2,\ldots \\ 0 & \text{en otro caso} \end{cases}$$

entonces X tiene función de densidad *Binomial Negativa* con parámetros *r* y *p* . Se denota como $X\sim B^{-}(r,p)$.

Ejemplo:

Supóngase que una persona juega poker hasta lograr tres flores imperiales (corrida del diez al as de la misma figura), ¿Cuál es la probabilidad de que se necesiten diez jugadas?.

La variable aleatoria X representa el número de jugadas en que no se logra realizar flores imperiales antes de que se obtengan 3. Entones: $r=3$, $x=7$, $p=20/52=0.38462$, $q=1-p=0.615538$ y

$$f_X(7)=\binom{9}{7}(0.38462)^3(0.61538)^7=0.00005,$$

por lo anterior la probabilidad de lograrlo es mínima.

vii) Ahora se define un caso particular de la función de densidad binomial negativa. Considérese una sucesión de n ensayos independientes de Bernoulli con igual probabilidad p de éxito. Sea X variable aleatoria que representa el número de intentos antes de obtener un éxito; entonces X tiene *función de densidad geométrica*,

$$f_X(x)=\begin{cases} p\,(1-p)^x & x=0,1,2,\dots \\ 0 & \text{en otro caso} \end{cases}$$

En el ejemplo anterior, del poker, ¿Cuál es la probabilidad de obtener un flor imperial en diez jugadas?.

$$f_X(10)=(0.38462)(0.61538)^{10}=0.0029.$$

VARIABLES ALEATORIAS CONTINUAS.

En el presente trabajo se asume que X, variable aleatoria continua, tiene un determinado comportamiento según convenga para la ilustración del modelo. Sin embargo, se puede utilizar la estadística descriptiva para observar el comportamiento del fenómeno y así aplicar el modelo que mejor lo describa.

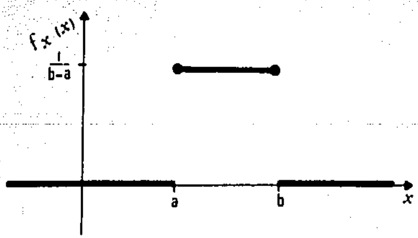
i) Sean *a* y b dos constantes, con $a<b$, se dice que *X* variable aleatoria continua tiene función de densidad *uniforme* en el intervalo (a,b) si la función de densidad esta dada por:

$$f_X(x) = \begin{cases} \dfrac{1}{b-a} & a<x<b \\ 0 & \text{en otro caso} \end{cases}$$

y su función de distribución es,

$$F_X(x) = \int_{\infty}^{x} f_X(t)\,dt = \begin{cases} 0 & x<a \\ (x-a)/(b-a) & a \le x \le b \\ 1 & x>b \end{cases}$$

Se denota como $X \sim U(a,b)$ y su gráfica es como sigue:

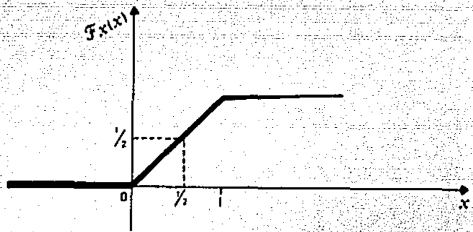


Ejemplo:

Supóngase que se tiene una circunferencia marcada del 0 al 1 y que por razones de simetría la probabilidad de que la manecilla se detenga en un intervalo es igual a la longitud del intervalo. ¿Cuál es la probabilidad de que la manecilla se detenga antes de 1/2?. Sea X el número en el que se detiene la manecilla y como $a=0$ y $b=1$, entonces

$$F_X(X) = \begin{cases} 0 & x<0 \\ x & 0\leq x<1 \\ 1 & \text{en otro caso} \end{cases}$$

Sea $x=1/2$, entonces $P(X\leq 1/2)=0.5$; su gráfica es como sigue:



ii) Una de las funciones de distribución más importante, por sus aplicaciones es la función de distribución normal; para introducirla considérese lo siguiente:

Sea $c = e^{-x^2/2}$ se desea que sea función de densidad; primero se necesita evaluarla,

$$c = \int_{-\infty}^{\infty} e^{-x^2/2} dx$$

se multiplica ambas partes por si misma

$$c^2 = \int_{-\infty}^{\infty} e^{-x^2/2}dx \int_{-\infty}^{\infty} e^{-y^2/2}dy$$

$$= \int_{-\infty}^{\infty} e^{-(x^2+y^2)/2}\, dxdy$$

sea $x=r\cos(\theta)$ $y=r\mathrm{sen}(\theta)$ entonces:

$$c^2 = \int_0^{\infty}\int_{-\pi}^{\pi} e^{-r^2/2} r d\theta\, dr$$

$$= 2\pi \int_0^{\infty} r\, e^{-r^2/2} dr$$

$$= -2\pi\, e^{-x^2/2} \Big|_0^{\infty} = 2\pi$$

por lo tanto $c = \sqrt{2\pi}$, lo que lleva a poder definir,

$$\phi_X(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \qquad -\infty < x < \infty$$

a esta función se le llama función de densidad *normal estándar*. Su función de distribución se denota como

$$\Phi_x(x) = \int_{-\infty}^{x} \phi_x(u)\,du.$$

La función de densidad $\phi_x(x)$ cumple:

a) $\phi_X(x)dx > 0$

b) $\int_{-\infty}^{\infty} \phi_X(x)dx = 1.$

Sea X variable aleatoria con función de densidad normal estándar y sea $Y = \mu + \sigma X$ donde $\sigma > 0$ entonces resulta

$$g_X(x) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp \left[\frac{-(X-\mu)^2}{2\sigma^2} \right] \quad \text{con } x \in \mathbb{R}$$

a esta función se le conoce como función de densidad *normal* con parámetros μ y σ^2. Se denota como $N(\mu,\sigma^2)$.

La función de distribución de $g_X(x)$ puede ser calculada en términos de la función Φ, es decir

$$\begin{aligned} P(Y \le y) &= P(\mu+\sigma X \le y) \\ &= P\left(X \le \frac{y - \mu}{\sigma} \right) \\ &= \Phi\left(\frac{y - \mu}{\sigma} \right). \end{aligned}$$

Dentro de las funciones de distribución existen algunas que son simétricas, es decir, que $F_X(-x)=1-F_X(x)$. Además se verifica que si $F_X(x)$ es continua, diferenciable, y simétrica entonces $f_X(-x)=f_X(x)$, es decir, $f_X(x)$ es simétrica con respecto al cero; el recíproco también se cumple, si $f_X(x)=f_X(-x)$, entonces $F_X(x)$ es simétrica,

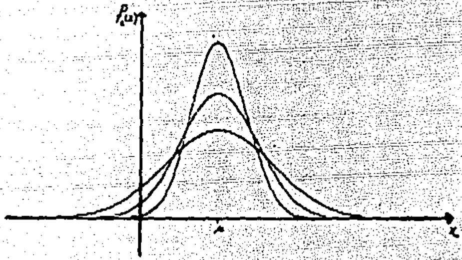
$$\begin{aligned} F_X(-x) &= \int_{-\infty}^{-x} f_Y(y)\,dy \\ &= \int_{x}^{\infty} f_Y(-y)\,dy \\ &= \int_{x}^{\infty} f_Y(y)\,dy \\ &= \int_{-\infty}^{\infty} f_Y(y)\,dy - \int_{-\infty}^{x} f_Y(y)\,dy \end{aligned}$$

por lo tanto

$$F_X(-x)=1-F_X(x) \qquad \text{para toda } -\infty<x<\infty.$$

En particular, la función de distribución de $\Phi_X(x)$ es simétrica, entonces $\Phi(-x) = 1 - \Phi(x)$.

Dependiendo de los valores de μ y σ^2 se obtienen diferentes curvas como se aprecia en las siguiente gráfica,



Ejemplo:

Supóngase que un profesor decide que la calificación final de cada alumno sea una variale aleatoria con distribución normal y parámetros (μ,σ^2). El maestro decide calificar con MB a los alumnos cuyas calificaciones excedan de $\mu+\sigma$, ¿Cuál es la probabilidad de obtener una MB?.

$$P(X>\mu+\sigma) = 1-P(\mu+\sigma) = 1-\Phi\left(\frac{\mu+\sigma-\mu}{\sigma}\right)$$
$$= 1-\Phi(1) \approx 0.1587,$$

por lo tanto la probabilidad de obtener MB es de 15.87%.

Una de las aplicaciones de la función de distribución Normal se analiza en el Capítulo 7, en donde se presenta el Teorema del Límite Central.

iii) Ahora, se analiza otra función de distribución, que también juega un papel importante dentro de la Teoría de Probabilidad, esta es la función de distribución Gamma. Para ello, defínase a la función Gamma como:

$$\Gamma(x)=\int_0^{\infty} e^{-t}t^{x-1}dt \qquad \text{con } x>0$$

Se desea que sea función de densidad, para ello, supóngase que $x>1$ entonces $x-1>0$, se integra por partes y resulta:

$$u=t^{x-1} \qquad du=(x-1)t^{x-2}dt$$
$$dv=e^{-t}dt \qquad v=-e^{-t}$$

entonces

$$\Gamma(x) = \int_0^{\infty} e^{-t}\, t^{x-1}\, dt = -t^{x-1}e^{-t}\Big|_0^{\infty} + \int_0^{\infty} (x-1)t^{x-2}\, e^{-t}dt$$

$$= (x-1)\int_0^{\infty} e^{-t}t^{x-2}dt$$

$$= (x-1)\Gamma(x-1),$$

integrando nuevamente por partes se tiene que

$$\Gamma(x) = (x-1)(x-2)\Gamma(x-2)$$

y así sucesivamente, entonces si x es un entero positivo

$$\Gamma(x) = (x-1)!.$$

Sea $\Gamma(\alpha,\lambda) = \int_0^\infty e^{-\lambda t} t^{\alpha-1} dt$ $\lambda>0, \alpha>0$

haciendo el cambio de variable $y=\lambda t$, $dy=\lambda dt$ resulta

$$\Gamma(\alpha,\lambda) = \int_0^\infty e^{-y} \left(\frac{y}{\lambda}\right)^{\alpha-1} \frac{dy}{\lambda}$$

$$= \frac{1}{\lambda^\alpha} \int_0^\infty e^{-y} y^{\alpha-1} dy$$

$$= \frac{\Gamma(\alpha)}{\lambda^\alpha}$$

por lo que finalmente, la función definida por:

$$\Gamma(\alpha,\lambda) = \begin{cases} \dfrac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} & x>0,\ \lambda>0,\ \alpha>0 \\ 0 & \text{en otro caso} \end{cases}$$

se llama la función de densidad *Gamma*, con parámetros (α,λ). Se verifica facilmente que $\Gamma(\alpha,\lambda)$, así, definida es función de densidad de probabilidad.

$$\int_0^\infty \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} dx = \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^\infty e^{-\lambda x} x^{\alpha-1} dx$$

$$= \frac{\lambda^\alpha}{\Gamma(\alpha)} \frac{\Gamma(\alpha)}{\lambda^\alpha} = 1. \quad \blacksquare$$

Ahora se enuncian algunos casos particulares de esta función de densidad:

a) En el caso de $\alpha=1$, $\lambda>0$ resulta que:

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & x>0,\ \lambda>0, \\ 0 & \text{en otro caso} \end{cases}$$

a esta función se llama *función de densidad Exponencial.*

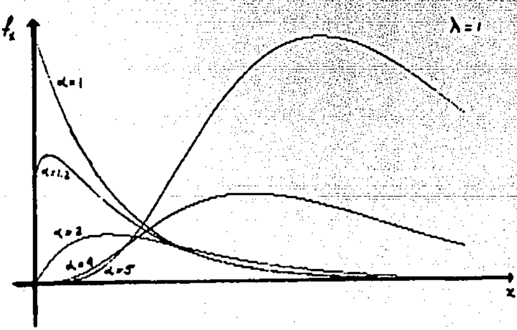
b) En el caso de que $\alpha = \frac{k}{2}$, $\lambda = \frac{1}{2}$, resulta,

$$f_X(x) = \frac{1}{\Gamma(k/2)} \left(\frac{1}{2}\right)^{k/2} x^{k/2-1} e^{-1/2x}$$

A esta función se le llama *función de densidad Chi-Cuadrada* con k grados de libertad. Se denota como $X^2(k)$.

Esta función permite generar otras funciones de distribución tales como la *F* ó la *t*-student, las cuales se estudian en los capítulos subsecuentes.

Dependiendo de los valores que α y λ, se obtienen diferentes curvas como se aprecia en la siguiente gráfica:



iv) Ahora, se analiza la función *Beta*, la cual se define por:

$$B(m,n)= \int_0^1 x^{m-1}(1-x)^{n-1}dx \qquad m,n>0$$

se desea que sea función de densidad, para ello primero se evalúa, por lo que se hace el siguiente cambio de variable,

$$\text{sea} \quad x = \frac{\tau}{1+\tau} \Rightarrow \tau = \frac{x}{1-x} \qquad 0 \le x \le 1$$

$$dx = \frac{1}{(1+\tau)^2}\, d\tau$$

entonces

$$B(m,n) = \int_0^\infty \frac{\left(\frac{\tau}{1+\tau}\right)^{m-1}\left(1-\frac{\tau}{1+\tau}\right)^{n-1}}{(1+\tau)^2}\, d\tau$$

$$= \int_0^\infty \frac{\tau^{m-1}}{(1+\tau)^{m-1}(1+\tau)^{n-1}(1+\tau)^2}\, d\tau$$

$$= \int_0^\infty \frac{\tau^{m-1}}{(1+\tau)^{m+n}}\, d\tau$$

por lo tanto

$$B(m,n) = \int_0^\infty \frac{\tau^{m-1}}{(1+\tau)^{m+n}}\, d\tau,$$

como se vió con la función *Gamma*

$$\int_0^\infty e^{-\lambda x}x^{n-1}dx = \frac{\Gamma(n)}{\lambda^n}$$

entonces

$$\Gamma(n) = \lambda^n \int_0^\infty e^{-\lambda t} t^{n-1} dt$$

y

$$\Gamma(m) = \int_0^\infty e^{-\lambda} \lambda^{m-1} d\lambda,$$

$$\Gamma(n)\Gamma(m) = \left[\lambda^n \int_0^\infty e^{-\lambda t} t^{n-1} dt\right] \left[\int_0^\infty \lambda^{m-1} e^{-\lambda} d\lambda\right]$$

$$= \int_0^\infty \left(\int_0^\infty \lambda^n e^{-\lambda t} t^{n-1} dt \right) \lambda^{m-1} e^{-\lambda} d\lambda$$

$$= \int_0^\infty \int_0^\infty \lambda^{m+n-1} e^{-\lambda t-\lambda} t^{n-1} dt d\lambda$$

$$= \int_0^\infty t^{n-1} \int_0^\infty \lambda^{m+n-1} e^{-(1+t)\lambda} d\lambda dt$$

$$= \int_0^\infty t^{n-1} \frac{\Gamma(m+n)}{(1+t)^{m+n}} dt$$

$$= \Gamma(m+n) \int_0^\infty \frac{t^{n-1}}{(1+t)^{m+n}} dt$$

$$= \Gamma(m+n) B(n,m)$$

por lo tanto

$$\frac{\Gamma(n)\Gamma(m)}{\Gamma(n+m)} = B(n,m),$$

por lo que finalmente, la función definida por:

$$B(n,m)=\begin{cases} \dfrac{\Gamma(m+n)\,x^{n-1}(1-x)^{m-1}}{\Gamma(m)\Gamma(n)} & x\in(0,1) \quad m,n>0 \\ 0 & \text{en otro caso} \end{cases}$$

se llama la función de densidad *Beta* con parámetros (n,m). Se verifica rápidamente que $B(n,m)$, así definida es función de densidad de probabilidad.

$$\int_0^\infty \frac{\Gamma(m+n)\,x^{n-1}(1-x)^{m-1}}{\Gamma(m)\Gamma(n)}\,dx = \frac{\Gamma(m+n)}{\Gamma(m)\Gamma(n)}\int_0^\infty x^{n-1}(1-x)^{m-1}dx$$

$$= \frac{1}{B(n,m)}\quad B(n,m) = 1. \quad \blacksquare$$

v) Sea X variable aleatoria continua, entonces se dice que X tiene función de densidad *Cauchy* si

$$f_X(x) = \frac{1}{\pi\,(1+x^2)} \qquad -\infty < x < \infty$$

y su función de distribución es

$$F_X(x) = \int_{-\infty}^{x} \frac{du}{\pi(1+u^2)}$$

$$= \frac{1}{2} + \frac{1}{\pi}\ \arctan x\ .$$

La función de distribución Cauchy tiene ciertas propiedades muy especiales; las cuales, se estudian en los capítulos IV y V.

vi) Una variable aleatoria continua X tiene función de densidad *Doble Exponencial* ó *Laplace* si

$$f_X(x) = \frac{1}{2\beta}\ \exp\left(\frac{|x-\alpha|}{\beta}\right) \qquad -\infty<\alpha<\infty \ \text{y}\ \beta>0\ .$$

VECTORES ALEATORIOS

Una variable aleatoria se había interpretado como una pregunta acerca del fenómeno aleatorio, sin embargo puede ocurrir que se tengan dos o más preguntas acerca del mismo fenómeno por lo que se obtiene un vector de variables aleatorias. Considérese que se quiere realizar un estudio sobre la fertilidad de la mujer mexicana, para ello se necesita seleccionar aleatoriamente a las mujeres que pesen menos de 50 Kg., tengan menos de 30 años, y un ingreso neto mensual menor de $2,000,000. En este caso el vector aleatorio (X,Y,Z) representaría las características de las mujeres.

Primero se analiza el caso de vectores aleatorios de dimensión 2, posteriormente se presenta la generalización de los conceptos a vectores de dimensión n. La siguiente definición formaliza el concepto de vector aleatorio.

<u>Definición 3.1.</u>

Sea $(\Omega,\mathbb{A},P)$ un espacio de probabilidad, X y Y dos variables aleatorias sobre el mismo espacio de probabilidad, esto es:

$X:\Omega\to\mathbb{R}$ tal que para toda $x\in\mathbb{R}$, $(w\in\Omega\,|\,X(w)\le x)\in\mathbb{A}$,
$Y:\Omega\to\mathbb{R}$ tal que para toda $y\in\mathbb{R}$, $(w\in\Omega\,|\,Y(w)\le y)\in\mathbb{A}$

entonces a (X,Y) se le llama un vector aleatorio de dimensión 2.

Ejemplo:

Sea c = (casamientos en el mes de noviembre en la Cd. de México),
$\Omega = (\mathbb{N})$,
X = contrayentes mayores de 18 años
Y = contrayentes extranjeros

entonces el vector (X,Y) representa a los contrayentes mayores de 18 años y extranjeros.

Una vez que se ha identificado el conjunto de posibles resultados de los fenómenos aleatorios surgen interrogantes acerca de él; por ejemplo sea

c = (lanzar dos monedas)
Ω = ((sol, sol), (águila, sol), (sol, águila), (águila, águila)).

Naturalmente, aparecen los siguientes cuestionamientos:

¿Obtendré dos águilas?,
¿Obtendré una águila y un sol?,
etc... .

Intuitivamente se observa que las preguntas asociadas a Ω son los compomentes de un vector aleatorio. Esto es, se puede asociar la obtención del resultado de lanzar la primera moneda a la primer componente del vector, y el resultado de la otra moneda a la segunda componente. El siguiente teorema formaliza este concepto.

Teorema 3.1.

Sea (Ω,A,P) un espacio de probabilidad, X y Y dos variables aleatorias. Entonces el subconjunto A tal que simultaneamente $X(w)\leq x$ y $Y(w)\leq y$ donde x y y son números reales, es un evento.

Demostración:

Sea $A=(w|X(w)<x$ y $Y(w)<y)$, es decir,

$$A=\left[(w|X(w)<x)\cap(w|Y(w)<y)\right]$$

como X y Y son dos variable aleatorias entonces $(w|X(w)<x)$ y $(w|Y(w)<y)$ son eventos; entonces A es la intersección de dos

eventos por la definición del σ-álgebra de eventos se deduce que A es un evento. ∎

Retomando el ejemplo de lanzar dos monedas si

X=(el resultado de la primera moneda sea águila),
Y=(el resultado de la segunda moneda sea sol)

entonces el conjunto de valores del vector $Z=(X,Y)$ es un evento.

Ahora bien, en el caso de que la dimensión del vector sea 2 o 3 puede ligarse a representaciones geométricas. En el caso de 2 variables aleatorias X y Y siempre se pueden considerar como coordenadas cartesianas rectangulares de un punto aleatorio en el plano. Análogamente, tres variables aleatorias se pueden considerar como coordenadas cartesianas rectangulares de un punto aleatorio en el espacio tridimensional. En el caso de que el vector aleatorio tenga una dimensión superior a 3 se comprenderá simplemente como coordenadas cartesianas de un punto n-dimensional aleatorio sin representación geométrica.

Dependiendo del fenómeno aleatorio, y de los valores que tome cada variable se tendran tres tipos de vectores aleatorios:

i) Discretos.- Si X y Y son variables aleatorias discretas
ii) Continuos.- Si X y Y son variables aleatorias continuas
iii) Mixtos.- Si X es variable aleatoria discreta y Y es variable aleatoria continua o viceversa.

En el presente trabajo sólo se analizan los vectores aleatorios del tipo *i)* y *ii)* debido a que se consideran la base para estudios formales de probabilidad.

FUNCION DE DISTRIBUCION ACUMULADA.

De igual forma, que en el caso de las variables aleatorias, lo que se desea es asignar probabilidades a los valores que toman los vectores aleatorios, para ello se introduce el concepto de función de distribución acumulada para vectores aleatorios.

Definición 3.2.

Una función $F:\mathbb{R}^2\to\mathbb{R}$ que cumple con:

i) $F(-\infty,-\infty)=0$

ii) $F(\infty,\infty)=1$

iii) F es no decreciente por variable

iv) F es continua por algún lado por variable (en particular se trabaja continua por la derecha)

v) Para toda $a\le b$, $c\le d$

$$F(b,d)-F(a,d)-F(b,c)+F(a,c)\ge 0$$

se le llama función distribución conjunta.

Proposición 3.1.

Sea $F:\mathbb{R}^2\to\mathbb{R}$ definida por

$$F_{XY}(x,y)=P(X\le x,\ Y\le y) \qquad -\infty<x,y<\infty$$

entonces F_{XY} así definida es la función de distribución conjunta de (X,Y).

Demostración.

i) Cuando $x\to-\infty$ y $y\to-\infty$ el cumplimiento de las desigualdades $X\le x$, $Y\le y$ tiende a un evento imposible por lo que $F(-\infty,-\infty)=0$.

ii) Cuando $x\to\infty$ y $y\to\infty$ el cumplimiento de la desigualdades $X\le x$, y $Y\le y$ tiende a un evento cierto, por lo que $F(\infty,\infty)=1$.

iii) Para y fijo se tiene que si $x_1 \le x_2 \Rightarrow F_{XY}(x_1,y) \le F_{XY}(x_2,y)$ y para x fija se tiene que si $y_1 \le y_2 \Rightarrow F_{XY}(x,y_1) \le F_{XY}(x,y_2)$, la demostración es igual a la presentada en el capítulo II.

iv) Supóngase y fija, si $x_n \to x$ por la derecha entonces

$$F_{XY}(x_n,y) \to F_{XY}(x,y)$$

en el caso de x fija si $y_n \to y$ por la derecha entonces

$$F_{XY}(x,y_n) \to F_{XY}(x,y)$$

por lo que F_{XY} es continua por la derecha por variable.

v) Sea $a \le b$ y $c \le d$. Entonces

$$P(a<X\le b, c<Y\le d)=F(b,d)-F(a,d)-F(b,c)+F(a,c)>0$$

Demostración.

$$1) \quad (X\le b, Y\le d)=(X\le a, Y\le c) \cup (X\le a, c<Y\le d) \cup (a<X\le b, Y\le c) \cup (a<X\le b, c<Y\le d),$$

$$2) \quad (X\le a, Y\le d)=(X\le a, Y\le c) \cup (X\le a, c<Y\le d),$$

$$3) \quad (X\le b, Y\le c)=(X\le a, Y\le c) \cup (a<X\le b, Y\le c),$$

entonces

$$F(b,d)=F(a,c)+P(X\le a, c<Y\le d) +P(a\le X\le b, Y\le c) +P(a<X\le b, c<Y<d),$$

$$F(a,d)=F(a,c)+P(X\le a, c<Y\le d),$$

$$F(b,c)=F(a,c)+P(a<X<b, Y\le c)$$

por lo tanto

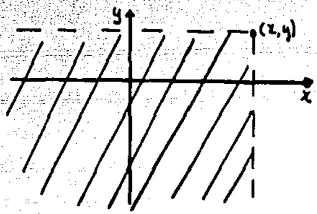
$$P(a<X\le b, c<Y\le d)=F(b,d)-F(a,c)-P(a\le X\le b, Y\le c)-P(a<X\le b, c<Y<d)$$

sustituyendo resulta

$$P(a<X\le b, c<Y\le d)=F(b,d)-F(a,c)+F(a,c)-F(a,d)-F(b,c)+F(a,c)$$

$$=F(b,d)-F(a,d)-F(b,c)+F(a,c)\ge 0. \quad \blacksquare$$

Geométricamente, la función de distribución de un vector bidimensional, representa la probabilidad de que un punto aleatorio en el plano, se encuentre a la izquierda y debajo del punto (x,y); se muestra como la parte sombreada en el diagrama.



Es claro que cada componente de un vector aleatorio es una variable aleatoria, por lo que tiene una función de distribución asociada. La función de distribución de cada componente se conoce como la función de distribución marginal. Este concepto se formaliza con la siguiente definición,

Definición 3.3.

Sea $F_{XY}(x,y)$ la función de distribución conjunta de X y Y, variables aleatorias entonces las funciones $F_X(x)$ y $F_Y(y)$ dadas por

$$F_X(x)=F(x,\infty)=\lim_{y\to\infty} F_{XY}(x,y)$$

$$F_Y(y)=F(\infty,y)=\lim_{x\to\infty} F_{XY}(x,y)$$

son llamadas funciones de distribución marginal de X y Y respectivamente.

La función de distribución marginal se puede interpretar como la función de distribución de x ó de y cuando la otra variable es ignorada.

FUNCION DE DENSIDAD CONJUNTA.

La pregunta natural que surge es ¿Cómo se calcula el valor de $f_{XY}(x,y)$?. Primero se analizará el caso de que el vector aleatorio (X,Y) sea discreto.

Definición 3.4.

Si (X,Y) es un vector aleatorio discreto, entonces la función de densidad conjunta esta definida por:

$$f_{XY}(x,y)=\begin{cases} P[X=x, Y=y] & \text{para } (x,y) \text{ valores de } (X,Y) \\ 0 & \text{en otro caso} \end{cases}$$

Está función cumple con las siguiente propiedades:

i) $f_{XY}(x,y) \geq 0$ por definición,

ii) $\Sigma f_{XY}(x,y) = 1$.

La función de distribución del vector (X,Y) se define como

$$F_{XY}(x,y)=\sum_{x}\sum_{y} f_{XY}(x,y).$$

En el caso de que (X,Y) sea un vector aleatorio continuo, lo natural sería que

$$F_{XY}(x,y)=\int_{-\infty}^{y}\int_{-\infty}^{x} f_{XY}(x,y)\,dxdy$$

y para que fuera cierto se necesitaría que

$$\frac{\delta^2}{\delta x \delta y} F_{XY}(x,y)=f_{XY}(x,y)$$

sin embargo, se conoce que tcda función continua por la derecha no necesariamente es derivable, entonces se necesita al igual que con las variables aleatorias, que el vector (X,Y) sea absolutamente continua.

<u>Definición</u> <u>3.5.</u>

Sea (X,Y) un vector aleatorio continuo, se dice que (X,Y) es absolutamente continua si existe una función $f:\mathbb{R}^2\to\mathbb{R}$ tal que f sea no negativa, y que

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XY}(x,y)\,dxdy=1$$

y para cualquier x, y$\in\mathbb{R}$

$$F_{XY}(x,y)=\int_{-\infty}^{y}\int_{-\infty}^{x} f_{XY}(x,y)\,dxdy$$

entonces a f_{XY} se le llama función de densidad del vector aleatorio (X,Y).

Función de Densidad Marginal.

Si se conoce la función de densidad conjunta de X y Y, entonces se puede determinar la función de densidad marginal f_Y ó f_X. En el caso de que X y Y sean variables aleatorias discretas, se obtiene sumando sobre x ó sumando sobre y, respectivamente, es decir

$$f_X(x) = \sum_y f_{XY}(x,y),$$

$$f_Y(y) = \sum_x f_{XY}(x,y).$$

Ejemplo:

Supóngase que de tres cartas numeradas del 1 al 3, dos son tomadas al azar. Sea X el número de la primera carta y Y el número de la segunda carta. Entonces la función de densidad f de (X, Y) se puede dar por $f(1,2)=f(1,3)=f(2,1)=f(2,3)=f(3,1)=f(3,2)=1/6$ y $f(x,y)=0$ en cualquier otro caso. La función de densidad marginal de X está dado por:

$$f_X(1)=f(1,1)+f(1,2)+f(1,3)=0+1/6+1/6=2/6=1/3.$$

En el caso de que X y Y sean variables aleatorias continuas resulta:

$$F_X(x)=P(X\leq x) = \int_{-\infty}^{x} \left(\int_{-\infty}^{\infty} f_{UY}(u,y)\,dydu\right)$$

entonces

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x,y)\,dy$$

y satisface que

$$F_X(x) = \int_{-\infty}^{x} f_X(u)\,du,$$

de la misma forma se obtiene $f_Y(y)=\int_{-\infty}^{\infty} f_{XY}(x,y)\,dx$.

Ejemplo:

Supóngase que se tiene a la función $f_{XY}(x,y)=x+y$ para $0<x<1$, $0<y<1$, ¿Es función de densidad?, ¿Cuál es $f_Y(y)$?.

i) $f_{XY}(x,y)=x+y\geq 0$

ii) $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XY}(x,y)\,dxdy= \int_{0}^{1}\int_{0}^{1}(x+y)\,dxdy$

$$= \int_0^1 (1/2 + y)\,dy$$

$= (1/2 + 1/2) = 1$, por lo tanto es función de densidad. Ahora

$$f_Y(y) = \int_0^1 (x+y)\,dx = (1/2 + y). \qquad \blacksquare$$

FUNCION DENSIDAD CONDICIONAL.

En la práctica es frecuente que se desee conocer la probabilidad de *X* a condición de que otra variable aleatoria *Y* ocurra o viceversa. De acuerdo con esto se determina la función de densidad condicional.

Definición 3.6.

Sea (X,Y) un vector aleatorio entonces, la función $f_{Y|X}$ definida por

$$f_{Y|X}(y|x) = \begin{cases} \dfrac{f_{XY}(x,y)}{f_X(x)} & f_X(x) > 0 \end{cases}$$

se le llama *función de densidad condicional* de *Y* dado *X*. Para cualquier valor de x.

Claramente se puede observa $f_{Y|X}(y|x)$ es función de densidad. En el caso de que (X,Y) sea un vector aleatorio discreto

$$\sum f_{Y|X}(y|x) = \frac{\sum_y f_{XY}(x,y)}{f_X(x)} = \frac{f_X(x)}{f_X(x)} = 1.$$

En el caso de que (X,Y) sea un vector aleatorio continuo, con función de densidad conjunta f_{XY},

$$\int_{-\infty}^{\infty} f_{Y|X}(y|x)\,dy = \int_{-\infty}^{\infty} \frac{f_{XY}(x,y)\,dy}{f_X(x)} = \frac{1}{f_X(x)} \int_{-\infty}^{\infty} f_{XY}(x,y)\,dy =$$

$$= \frac{f_X(x)}{f_X(x)} = 1.$$

Ejemplo:

Sea (X,Y) un vector aleatorio distribuido uniformemente dentro de un rectángulo con función de densidad

$$f_{XY}(x,y) = \begin{cases} 1/4\,ab & \text{si } |x| \le a,\ |y| \le b \\ 0 & \text{si } |x| > a,\ |y| > b \end{cases}$$

¿Cuál es $f_{Y|X}f(y|x)$?.

Primero se debe de determinar las funciones de densidad marginal

$$f_X(x) = \begin{cases} 1/2a & \text{si } |x| \le a \\ 0 & \text{si } |x| > a \end{cases}$$

y

$$f_Y(y) = \begin{cases} 1/2b & \text{si } |y| \le b \\ 0 & \text{si } |y| > b \end{cases}$$

entonces

$$f_{Y|X}(y|x) = \begin{cases} \dfrac{f_{XY}(x,y)}{f_X(x)} & 0 < f_x(x). \end{cases}$$

sustituyendo resulta que:

$$f_{Y|X}(y|x) = \begin{cases} \dfrac{f_{XY}(x,y)}{f_X(x)} = \dfrac{1}{2} & \text{si } |x|<a,\ |y|<b. \end{cases}$$

Ahora se define la Fórmula de Bayes para vectores aleatorios. Se ha definido a la probabilidad condicional como:

$$f_{Y|X}(x|y)=\frac{f_{XY}(x,y)}{f_Y(y)} \qquad 0<f_Y(y)$$

entonces

$$f_{XY}(x,y)=f_X(x)f_{Y|X}(y|x)$$

por otro lado se tiene

$$f_Y(y)=\int_{-\infty}^{\infty} f_{XY}(x,y)\,dx = \int_{-\infty}^{\infty} f_X(x)f_{Y|X}(y|x)\,dx$$

entonces la probabilidad condicional se puede definir como

$$f_{X|Y}(x|y)=\frac{f_X(x)f_{Y|X}(y|x)}{\int_{-\infty}^{\infty} f_X(x)f_{Y|X}(y|x)\,dx}$$

esta fórmula es análoga a la presentada en el capitulo I y se le conoce como la *Formula de Bayes* para vectores aleatorios.

INDEPENDENCIA

Si a una persona se le hiciera corresponder su talla de ropa con el número de habitaciones en su casa; hay razón para pensar que no existe ninguna relación entre estas dos variables ya que el hecho de tener una determinada talla no influye para nada con el número de habitaciones en la casa de alguien. Siempre se encontraran variables aleatorias que no tienen ninguna influencia entre sí, por lo que se dirá que son independientes; la siguiente definición formaliza el concepto de independencia.

Definición 3.7.

Se dice que dos variables aleatorias con función de distribución conjunta son independientes si

$$F_{XY}(x,y)=F_X(x)F_Y(y) \qquad -\infty<x,y<\infty.$$

Definición 3.8.

Se dice que dos variables aleatorias con función de densidad conjunta son independientes si

$$f_{XY}(x,y)=f_X(x)f_Y(y) \qquad -\infty<x,y<\infty.$$

Existe una gran relación entre el concepto de independencia y el de probabilidad condicional. Supóngase que X y Y son dos variables aleatorias independientes entonces

$$f_{XY}(x,y)=f_X(x)f_Y(y)$$

por definición de densidad condicional se tiene que

$$f_{XY}(x,y)=f_{Y|X}(y|x)f_X(x)$$

lo cual implica que

$$f_{Y|X}(y|x)=f_Y(y),$$

es decir, dos variables son independientes cuando el conocimiento del valor de una de ellas no varia la ley de distribución de la otra. Con otras palabras, la información que se obtiene fijando el valor de una de las variables aleatorias no contiene ninguna injerencia sobre la segunda variable. Por lo tanto, para demostrar que dos variables son independientes, basta con mostrar que $f_{y|x}(y|x)$ no depende de x.

Ejemplo:

Supóngase que el tiempo que toma a 2 estudiantes resolver un problema es independiente y distribuido de manera exponencial con parámetro λ. Encontrar la probabilidad de que el primer estudiante tome al menos el doble del tiempo que el segundo estudiante en resolver el problema.

Sea X=tiempo que toma al primer estudiante resolver el problema,
Y=tiempo que toma al segundo estudiante resolver el problema;

ya que el tiempo que tardan se distribuye exponencialmente resulta que

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & \text{en otro caso} \end{cases}$$

de la misma forma resulta $f_y(y)$, y como son independientes entonces la función de densidad conjunta es:

$$f_Y(y) f_X(x) = f_{XY}(x,y) = \begin{cases} \lambda^2 e^{-\lambda x} e^{-\lambda y} & x,y \geq 0 \\ 0 & \text{en otro caso} \end{cases}$$

entonces

$$P(X \geq 2Y) = \int_0^{\infty} \int_0^{x/2} f_{XY}(x,y)\,dxdy = \int_0^{\infty} \int_0^{x/2} \lambda^2 e^{-\lambda x} e^{-\lambda y}\,dxdy$$

$$= \lambda^2 \int_0^\infty e^{-\lambda x} \int_0^{x/2} e^{-\lambda y} dy dx$$

$$= \frac{-\lambda^2}{\lambda} \int_0^\infty e^{-\lambda x} \, e^{-\lambda y} \Big|_0^{x/2} dx$$

$$= -\lambda \int_0^\infty e^{-\lambda x} (e^{-\lambda x/2} - 1) dx$$

$$= -\lambda \int_0^\infty (e^{-3/2\, \lambda x} - e^{-\lambda x}) dx = 1/3,$$

por lo tanto la probabilidad de que el primer estudiante tome al menos el doble del tiempo que el segundo es 1/3.

FUNCION DE DISTRIBUCION NORMAL BIVARIADA.

Debido a la importancia de la función de densidad normal, a continuación se define a la función de densidad normal bivariada, es decir para (X,Y) variables aleatorias, asimismo se define su función de densidad marginal y su función de densidad condicional.

Definición 3.9.

Sea (X,Y) dos variables aleatorias con función de densidad conjunta, entonces la función de densidad normal bivariada se define como:

$$f_{XY}(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}}$$

$$\exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_x)^2}{\sigma_x^2} - \frac{2\rho}{\sigma_x\sigma_y}(x-\mu_x)(y-\mu_y) + \frac{(y-\mu_y)^2}{\sigma_y^2}\right]\right\}$$

para $(x,y)\in\mathbb{R}^2$ $(\mu_x.\mu_y)\in\mathbb{R}^2$ $\sigma_x,\sigma_y>0$ $-1<\rho<1$.

Ahora se tiene que comprobar que realmente es función de densidad conjunta es decir:

i) $f_{xy}(x,y)\geq 0$; el pimer miembro es positivo debido a que $\sigma_x,\sigma_y>0$ y como $-1<\rho<1$ implica que $0<1-\rho^2<1$ y como $\exp(x)>0$, entonces se cumple que $f_{xy}(x,y)\geq 0$.

ii) $\int_{-\infty}^{\infty}\int_{-\infty} f_{XY}(x,y)\,dxdy=1$.

Demostración:

Se realiza el siguiente cambio de variable

$$u=\frac{x-\mu_x}{\sigma x} \quad y \quad v=\frac{y-\mu_y}{\sigma_y}$$

entonces queda

$$f_{XY}(x,y)=\int_{-\infty}^{\infty}\int_{-\infty}\frac{1}{2\Pi\sqrt{1-\rho^2}}\exp\left\{-\frac{1}{2(1-\rho^2)}\left[u^2-2\rho uv+v^2\right]\right\}dudv$$

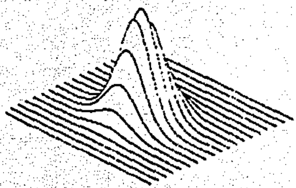
completando cuadrados resulta que:

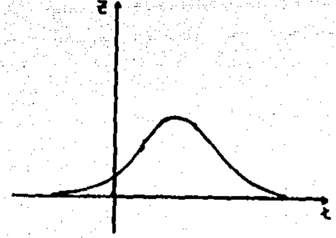
$$u^2-2\rho uv+v^2=(u-\rho v)^2-\rho^2v^2+v^2=(u-\rho v)^2+v^2(1-\rho^2)$$

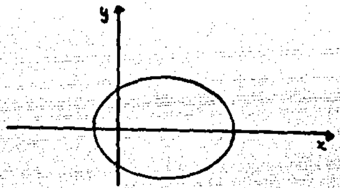
$$f_{XY}(x,y)=\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-v^2/2}dv\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}\sqrt{1-\rho^2}}e^{-(u-\rho v)^2/2(1-\rho^2)}du$$

$$=1.$$

por lo que se establece que es función de densidad conjunta y es llamada *Normal Bivariada*. En la siguiente figura se aprecia su gráfica, asicomo sus cortes.







Cualquier plano paralelo al plano xy que corte la superficie intersectará con una curva elíptica; cualquier plano perpendicular al plano xy intersectará con una curva normal. La función de densidad marginal determina la función de distribución de estos cortes.

Ahora se analiza las funciones de densidad marginal. Por definición de Densidad Marginal se tiene que:

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x,y)\,dy,$$

se realiza un cambio de variables $v = \dfrac{y-\mu_y}{\sigma_y}$ y $dv = \dfrac{dy}{\sigma_y}$
entonces

$$f_X(x) = \int_{-\infty}^{\infty} \frac{1}{2\pi\sigma_x\sqrt{1-\rho^2}}$$

$$\exp\left\{\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu x)^2}{\sigma x^2}-\frac{2\rho}{\sigma x}(x-\mu x)v+v^2\right]\right\}dv$$

y completando el cuadrado de v se tiene que

$$f_X(x)=\left(\int_{-\infty}^{\infty}\frac{1}{2\pi\sigma_x\sqrt{1-\rho^2}}\right.$$

$$\left.\exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(v-\rho\frac{(x-\mu_x)}{\sigma_x}\right)^2+(1-\rho^2)\frac{(x-\mu_x)^2}{\sigma x^2}\right]\right\}dv\right)$$

$$=\frac{1}{\sqrt{2\pi}\ \sigma x}e^{-(x-\mu_x)^2/2\sigma x^2}$$

$$\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}\sqrt{1-\rho^2}}\exp\left\{\frac{1}{2(1-\rho^2)}\left(v-\rho\frac{(x-\mu x)^2}{\sigma x^2}\right)\right\}dv=$$

$$=\frac{1}{\sqrt{2\pi}\ \sigma_x}e^{-(x-\mu x)^2/2\sigma x^2}\sim N(\mu x,\sigma x^2);$$

de la misma manera se puede obtener $f_Y(y)$,

$$f_Y(y)=\frac{1}{\sqrt{2\pi}\ \sigma_y}e^{-(y-\mu y)^2/2\sigma y^2}\sim N(\mu_y,\sigma_y^2)$$

entonces si (X,Y) tiene función de densidad normal, las funciones de densidad marginal respectivamente también se distribuyen normalmente.

Función de Densidad Condicional de la Normal Bivariada.

Si en la función de densidad condicional, definida como

$$f_{X|Y}(x|y)=\frac{f_{XY}(x,y)}{f_Y(y)}$$

se sustituye resulta lo siguiente:

$$f_{X|Y}(x|y)=\frac{1}{\sqrt{2\pi}\,\sigma x\sqrt{1-\rho^2}}\exp\left\{-\frac{1}{2\sigma_x^2(1-\rho^2)}\left[x-\mu_x-\frac{\rho\sigma x}{\sigma y}(y-\mu_y)\right]^2\right\},$$

i) si $\rho=0$ se tiene

$$f_{X|Y}(x|y)=\frac{1}{\sqrt{2\pi}\,\sigma x}\exp\left\{-\frac{1}{2\sigma_x^2}(x-\mu_x)^2\right\}\sim N(\mu_x,\sigma x^2)$$

y se puede observar que $f_{X|Y}(x|y)$ no depende de y, por lo que X y Y son independientes.

ii) si $\rho\neq 0$ se ve claramente que X y Y no son independientes.

VECTORES ALEATORIOS DE DIMENSION N.

Como se mencionó anteriormente puede ocurrir que en el estudio de los fenómenos aleatorios se tengan 2 ó más preguntas acerca del fenómeno, por lo que en caso de tener n preguntas se trabaja con vectores aleatorios de dimensión n.

Algunas demostraciones presentadas para vectores bidimensionales, son análogas para vectores de dimensión n, por lo que se consideran como válidas y no se realizarán nuevamente.

Definición 3.10.

Si $X_1,X_2,X_3,\ldots,X_n$ son variables aleatorias definidas en el mismo espacio de probabilidad $(\Omega,\mathbb{A},P)$ entonces al vector $(X_1,X_2,\ldots,X_n)$ se le llama vector aleatorio de dimensión n.

Dependiendo de los valores que tome cada X_i $i=1,..n$ el vector aleatorio se clasificará en:

i) Discreto.- Si $X_1,...,X_n$ son variables aleatorias discretas,
ii) Continuo.- Si $X_1,...,X_n$ son variables aleatorias continuas.

FUNCION DE DISTRIBUCION ACUMULADA.

Si $X_1,X_2,...,X_n$ son variables aleatorias definidas en un mismo espacio de probabilidad (Ω,A,P) entonces la función de distribución conjunta F se define:

$$F(x_1,x_2,...,x_n)=P(X_1\le x_1,X_2\le x_2,...,X_n\le x_n) \qquad -\infty<x_1,x_2,...,x_n<\infty$$

Las propiedades de la función de distribución deducidas anteriormente para los vectores aleatorios bidimensionales, se extienden a los vectores de un número cualesquiera de dimensión.

i) $F(x_1,x_2,...,x_n)=0$ si al menos uno de los x_i's $=-\infty$ y $F(\infty,\infty,...,\infty)=1$,
ii) $F(x_1,x_2,...,x_n)$ es no decreciente en cada uno de los x_i's,
iii) $F(x_1,x_2,...,x_n)$ es continua no decreciente por la derecha en cada una de las x_i's,
iv) Para toda $a_1\le b_1$, $a_2\le b_2$, $a_3\le b_3,...,a_n\le b_n$
$F(b_1,b_2,...,b_n)-F(a_1,b_2,...b_n)-F(b_1,a_2,b_3,...,b_n)-...$
$-F(b_1,b_2,...,b_{n-1},a_n)+$(todos los términos con 2 a's y n-2 b's)
-(todos los términos con 3 a's y n-3 b's)+...
$+(-1)^nF(a_1,a_2,...,a_n)\ge 0$,
en el caso bidimensional se demostró que:
$P(a_1<X\le b_1,a_2<Y\le b_2)=F(b_1,b_2)-F(b_1,a_2)-F(a_1,b_2)+F(a_1,a_2)>0$,
equivalentemente la propiedad *iv)* es:

$$P(a_1<X_1\le b_1,a_2<X_2\le b_2,...,a_n<X_n\le b_n)\ge 0.$$

Cada componente de un vector aleatorio es una variable aleatoria por lo que cada una tiene una función de distribución asociada, es decir

$$F_1(x_1)=P(X_1 \le x_1)$$

en este caso a $F_1(x_1)$ es la función de distribución marginal de X_1.

Uno de los objetivos de la teoría de probabilidad es cuantificar el valor de $F(x_1,x_2,...x_n)$; por lo cual se iniciará con vectores aleatorios discretos.

Caso Discreto.

En el caso de un vector aleatorio discreto la función de distribución acumulada multivariada es determinada por:

$$F(x_1,x_2,...,x_n)=\sum_{t_1 \le x_1} \cdots \sum_{t_n \le x_n} f(t_1,t_2,...,t_n)$$

en donde $f(x_1,x_2,...,x_n)$ es la función de densidad de $(X_1,X_2,...X_n)$. Al igual que el caso bidimensional, las propiedades de la función de densidad se extienden a vectores de un número cualquiera de dimensión, es decir

i) $\sum f(x_1,x_2,...,x_n)=1$ cuando la sumatoria es realizada sobre todos los posibles valores de $(X_1,X_2,...,X_n)$,

ii) $f(x_1,x_2,...,x_n)=P[X_1=x_1;X_2=x_2;...;X_n=x_n]$ para $(x_1,x_2,...x_n)$ valores de $(X_1,X_2,...,X_n)$ y 0 en cualquier otro caso.

Caso Continuo.

En el caso de un vector aleatorio continuo, se dice que $f(x_1,x_2,...,x_n)$ es función de densidad del vector aleatorio si cumple con:

$$\frac{\delta^n}{\delta x_1 \delta x_2 \ldots \delta x_n} F(x_1, x_2, \ldots) = f(x_1, x_2, \ldots x_n)$$

lo cual implica que la función de distribución es absolutamente continua y

$$F(x_1, x_2, \ldots, x_n) = \int_{-\infty}^{x_n} \ldots \int_{-\infty}^{x_1} f(x_1, x_2, \ldots x_n)\, dx_1 dx_2 \ldots dx_n$$

como se ha presentado anteriormente la función de densidad tiene las siguientes propiedades:

i) $f(x_1, x_2, \ldots, x_n) \geq 0$,

ii) $\int_{-\infty}^{\infty} \int_{-\infty} f(x_1, x_2, x_3, \ldots, x_n)\, dx_1 dx_2 \ldots d_{x_n} = 1.$

Sea $(X_{i1}, \ldots, X_{ik})$ un subconjunto de las variables aleatorias $(X_1, \ldots, X_m)$, $(m \leq k)$, entonces $f(x_{i1}, \ldots, x_{ik})$ es llamada la función de densidad marginal de dimensión k de la variable aleatoria $(X_{i1}, \ldots, X_{ik})$.

FUNCION DE DENSIDAD CONDICIONAL.

De la misma forma que para el vector (X,Y) aleatorio, puede surgir la necesidad de hallar la probabilidad de una serie variables aleatorias a condición de que ocurran otras variables aleatorias.

Definición 3.11.

Sea $(X_1, \ldots, X_k)$ un vector aleatorio de dimensión k y sea $X_{i1}, \ldots, X_{ir}$ y $X_{j1}, \ldots, X_{js}$ dos subconjuntos de variables aleatorias de $X_1, \ldots X_k$. La función de densidad condicional de la variable aleatoria $(X_{i1}, .. X_{ir})$ dado el valor $(x_{j1}, \ldots, x_{js})$ de $(X_{j1}, \ldots, X_{js})$ se determina por:

$$f_{X_{i1},\ldots,X_{ir}|X_{j1},\ldots,X_{js}}(x_{i1},\ldots,x_{ir}|x_{j1},\ldots,x_{js}) =$$

$$= \frac{f(x_{i1},\ldots,x_{ir},x_{j1},\ldots,x_{js})}{f(x_{j1},\ldots,x_{js})}$$

Ejemplo:

Supóngase que son repartidas 12 cartas al azar sin reemplazo. Sea X_1 el número de ace's, X_2 el número de 2's, X_3 el número de 3's y X_4 el número de 4's. La función de densidad está dada por:

$$f(x_1,x_2,x_3,x_4) = \frac{\binom{4}{x_1}\binom{4}{x_2}\binom{4}{x_3}\binom{4}{x_4}\binom{36}{12-x_1-x_2-x_3-x_4}}{\binom{52}{12}}$$

$$\text{y } f_{X_2X_4|X_1X_3}(x_2,x_4|x_1,x_3) = \frac{f(x_1,x_2,x_3,x_4)}{f(x_1,x_3)}$$

$$= \frac{\binom{4}{x_2}\binom{4}{x_4}\binom{36}{12-x_1-x_2-x_3-x_4}}{\binom{44}{12-x_1-x_3}}$$

INDEPENDENCIA.

El concepto de Independencia guarda su significado para los vectores aleatorios. Se dice que las variables aleatorias $X_1,\ldots,X_n$ son independientes si dos vectores cualesquiera, que se puedan formar de estas variables aleatorias y que no contienen elementos comunes son independientes, es decir, las variables aleatorias $X_1,\ldots,X_n$ son independientes cuando, y solamente cuando, cada magnitud X_i es independiente de las demás magnitudes $X_1,\ldots X_{i-1},X_i,\ldots X_n$ tomadas por separado, dos a dos, tres a tres,etc. De acuerdo a la definición de independencia resulta lo siguiente:

Definición 3.12.

Sea $(X_1, X_2, \ldots, X_n)$ un vector aleatorio de dimensión n, se dice que es independiente si:

$$F(x_1, \ldots, x_n) = \prod_{i=1}^{n} F(x_i).$$

Definición 3.13.

Sea $(X_1, X_2, \ldots, X_n)$ un vector aleatorio de dimensión n, se dice que es independiente si:

$$f(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} f(x_i).$$

Se ha observado que el concepto de independencia está intimamente ligado al concepto de distribución condicional, en el caso de vectores de variables aleatorios sigue guardando esa misma relación.

Supóngase $X_{i1}, \ldots, X_{ir}$ y $X_{j1}, \ldots, X_{js}$ dos subconjuntos independientes de variables aleatorias de $X_1, \ldots, X_k$ entonces

$$f(x_{i1}, \ldots, x_{js}) = \prod_{k=i1}^{js} f(x_k)$$

por definición de densidad condicional se tiene que

$$f_{x_{i1}, \ldots, {ir} | x_{j1}, \ldots, x_{js}}(x_{i1}, \ldots, x_{ir} | x_{j1}, \ldots, x_{js}) = f(x_{i1}, \ldots, x_{ir})$$

por lo tanto para demostrar que n variables son independientes, basta con mostrar que

$f_{x_{i1}, \ldots, x_{ir} | x_{j1}, \ldots, x_{js}}(x_{i1}, \ldots, x_{ir} | x_{j1}, \ldots, x_{js})$ no depende de $(x_{j1}, \ldots, x_{js})$.

Ejemplos de Funciones de Distribución Multivariada.

Caso Discreto.

Ejemplo:

Supóngase que se arroja un dado seis veces, sea E_i el evento de que la i-ésima cara resulte. Resulta que $p_i=1/6$, $i=1,2,...,6$. ¿Cuál es la probabilidad de obtener dos veces el número 1, una vez el número 2, el número 3, el número 5, el número 6 y ninguna vez el número 4?.

Para poder determinar la probabilidad antes señalada considérese lo siguiente:

N= el número de repeticiones independientes del experimento,

X_i=el número de veces que el evento E_i ocurre,

$\sum_{i=1}^{n} p_i=1$,

entonces

$$P(X_1=x_1, X_2=x_2, ..., X_n=x_n)=kp_1^{x_1}p_2^{x_2}...p_n^{x_n}$$

donde k es el número de arreglos posibles de N repeticiones en $x_1, x_2, ...x_n$, es decir, la probabilidad de ocurrencia del evento E_i en cada intento es p_i, entonces en N repeticiones el evento E_1 ocurre x_1 veces, el evento E_2, x_2 veces y así sucesivamente el evento E_n, x_n veces $(x_1+x_2+...+x_n=N)$,

$$k=\binom{N}{x_1, x_2 ... x_{n-1}}$$

por lo tanto

$$p(x_1, x_2, ..., x_n)=\binom{N}{x_1, x_2, ... x_{n-1}}p_1^{x_1}p_2^{x_2}...p_n^{x_n}$$

$$=\frac{N!}{x_1!x_2!...x_n!}p_1^{x_1}p_2^{x_2}...p_n^{x_n}$$

que es conocida como la Función de Distribución Multinomial. La Función de Distribución Acumulada está dada por:

$$F(x_1, x_2, \ldots x_n) = \sum_{t_1 \le x_1, \ldots, t_n \le x_n} \binom{N}{t_1, t_2, \ldots, t_{n-1}} p_1^{t_1} p_2^{t_2} \ldots p_n^{t_n}.$$

regresando al ejemplo resulta que:

$$p(2,1,1,0,1,1) = \frac{6!}{2!1!1!0!1!1!} \left(\frac{1}{6}\right)^6 = 0.0077.$$

Para evaluar la función de densidad marginal se realiza lo siguiente:

$$p_{1,2,\ldots,k}(x_1, x_2, \ldots x_n) =$$

$$= \sum_{x_{k+1}, x_{k+2}, \ldots, x_n} \binom{N}{x_1, x_2, \ldots, x_{n-1}} p_1^{x_1} p_2^{x_2} \ldots p_n^{x_n}$$

$$= \sum_{x_{k+1}, x_{k+2}, \ldots, x_n} \frac{N!}{x_1! x_2! \ldots x_n!} p_1^{x_1} p_2^{x_2} \ldots p_n^{x_n}$$

$$= \frac{N!}{x_1! x_2! \ldots x_k! \, (N - \sum_{j=1}^{k} x_j)!} p_1^{x_1} p_2^{x_2} \ldots p_k^{x_k} \left(1 - \sum_{j=1}^{k} p_j\right)^{N - \sum_{j=1}^{k} x_j}$$

$$\sum_{x_{k+1}, \ldots x_n} \frac{(N - \sum_{j=1}^{k} x_j)}{x_{k+1}! x_{k+2}! \ldots x_n!} \left(\frac{p_{k+1}}{q}\right)^{x_{k+1}} \cdots \left(\frac{p_n}{q}\right)^{x_n},$$

la sumatoria es igual a 1, debido a que se trata de la función de distribución multinomial. La $q>0$ puesto que no tendria sentido que $q=0$, por lo tanto

$$p_{1,2,\ldots k}(x_1, x_2, \ldots, x_k) =$$

$$= \frac{N!}{x_1! x_2! \ldots x_k! \, (N - \sum_{j=1}^{k} x_j)!} p_1^{x_1} p_2^{x_2} \ldots p_k^{x_k} \left(1 - \sum_{j=1}^{k} p_j\right)^{N - \sum_{j=1}^{k} x_j}$$

Para introducir el concepto de Transformación de variables aleatorias, considérense los siguientes ejemplos. Supóngase que la probabilidad de que una partícula radioactiva se desintegre durante un tiempo x, x>0, está dado por $P(X\leq x)=1-e^{-\alpha x}$, donde α es una constante; un investigador observa el tiempo de desintegración de un gran número de partículas y anota el logaritmo del tiempo, más que el tiempo en sí; por lo tanto si se desea conocer si el experimento confirma la teoría, se tiene que evaluar $P(\log X\leq y)$, tomando como base lo concerniente a la distribución de la variable aleatoria X. Pero ¿Como se determinaría la función de distribución de la variable aleatoria Y?. O bien supóngase que se desea encontrar la función de distribución del cociente de dos variables aleatorias independientes X_1, X_2 cada una con distribución X^2 y divididas por su respectivos grados de libertad, es decir $Y=\dfrac{X_1\ n_2}{X_2\ n_1}$, ¿Como se determinaría la distribución de Y?.

En los ejemplos anteriores, se tiene a una variable aleatoria que es transformada por alguna función $g(.)$, para definir una nueva variable aleatoria Y; en el primer ejemplo se tiene a la variable X que es transformada por la función logaritmo tal que define a $Y=\log X$, en el segundo ejemplo se tiene a la división que está transformando a X_1 y X_2 para definir a $Y=\dfrac{n_2\ X_1}{n_1\ X_2}$. Se trata de obtener la función de distribución de Y, a partir de las funciones de distribución que se conocen y de la transformación $g(.)$. La pregunta que surge es ¿Que condiciones debe cumplir $g(.)$ para que $g(X)=Y$ sea variable aleatoria?. Se analiza el caso de que $g(.)$ es una función continua.

Teorema 4.1.

Si X es una variable aleatoria continua con función de distribución $F_X(x)$ y $Y=g(X)$ entonces la función de distribución de Y se define como

$$F_Y(y) = \int_{g^{-1}(T_y)} dF_X(x)$$

donde $T_y=\{Y:Y\leq y\}$.

Demostración.

Sea $T_y=\{Y:Y\leq y\}$ y $F_Y(y)$ la función de distribución de Y, entonces $F_Y(y)=P(Y\leq y)=P(Y\in T)$ pero como $Y=g(X)$ resulta que

$$P(Y\in T)=P(X\in g^{-1}(T))=\int_{g^{-1}(T)} dF_X(x)$$

donde $g^{-1}(T)=\{x:g(x)\in T\}$. ∎

Proposición 1.

En el caso de que $g(.)$ sea una función monotona creciente y diferenciable, la función de distribución resulta ser:

$$F_Y(y) = P(g(X)\leq y) = P(X\leq g^{-1}(y)) = F_x(g^{-1}(y))$$

y la función de densidad

$$f_Y(y) = F_Y'(y) = F_x'(g^{-1}(y)\ (g^{-1}(y))' = f_x(g^{-1}(y))\ (g^{-1}(y))'.$$

Proposición 2.

En el caso de que $g(.)$ sea una función monotona decreciente y diferenciable la función de distribución resulta

$$F_Y(y) = P(g(x)\leq y) = P(x\geq g^{-1}(y)) = 1 - P(X\leq g^{-1}(y)) = 1 - F_x(g^{-1}(y))$$

y la función de densidad

$$f_Y(y) = F_Y'(y) = (1 - F_X(g^{-1}(y)))' = -F_X'(g^{-1}(y))(g^{-1}(y))'$$
$$= -f_x(g^{-1}(y))(g^{-1}(y))'.$$

Proposición 3.

Si $Y=g(X)=aX+b$, entonces para $a\neq 0$ resulta que si $a>0$

$$F_Y(y)=F_X\left(\frac{y-b}{a}\right)$$

si $a<0$

$$F_Y(y) = 1 - F_X\left(\frac{y-b}{a}\right)$$

y si $a=0$

$$F_Y(y)=\begin{cases} 0, & y<b \\ 1, & y\geq b \end{cases}$$

y la función de densidad es

$$f_Y(y)= \frac{1}{|a|}f_x\left(\frac{y-b}{a}\right)$$

la demostración es inmediata de los resultados obtenidos.

Proposición 4.

Si $Y=g(X)=aX^2+bX+c$ entonces para $a>0$

$$F_Y(y)=F_X\left(\frac{-b + \sqrt{b^2 - 4a(c-y)}}{2a}\right) - F_X\left(\frac{-b - \sqrt{b^2 - 4a(c-y)}}{2a}\right)$$

y si $a<0$

$$F_Y(y) = 1-F\left(\frac{-b+\sqrt{b^2-4a(c-y)}}{2a}\right)+F\left(\frac{-b-\sqrt{b^2-4a(c-y)}}{2a}\right),$$

la función de densidad queda definida como:

$$f_Y(y)=\frac{1}{\sqrt{b^2-4a(c-y)}}\left\{f_X\left(\frac{-b+\sqrt{b^2-4a(c-y)}}{2a}\right)+\right.$$

$$\left.f_Y\left(\frac{-b-\sqrt{b^2-4a(c-y)}}{2a}\right)\right\}.$$

Demostración:

Considérese para $a>0$

$$F_Y(y)=P(g(X)\leq y)=P(X\leq g^{-1}(y))=$$

$$=P\left(x|x\leq\frac{-b+\sqrt{b^2-4a(c-y)}}{2a}\text{ ó }x\geq\frac{-b-\sqrt{b^2-4a(c-y)}}{2a}\right)$$

$$=F_X\left(\frac{-b+\sqrt{b^2-4a(c-y)}}{2a}\right)-F_X\left(\frac{-b-\sqrt{b^2-4a(c-y)}}{2a}\right)$$

en el caso de que $b^2-4a(c-y)$ sea negativo entonces $F_Y(y)$ es cero. En el caso de $a<0$ resulta que

$$g^{-1}(y)=\left(x|x\leq\frac{-b+\sqrt{b^2-4a(c-y)}}{2a}\text{ ó }x\geq\frac{-b-\sqrt{b^2-4a(c-y)}}{2a}\right)$$

entonces

$$F_Y(y) = 1-F\left(\frac{-b + \sqrt{b^2 -4a(c-y)}}{2a}\right) +F\left(\frac{-b - \sqrt{b^2 -4a(c-y)}}{2a}\right)$$

derivando resulta que

$$f_Y(y) = f_X\left(\frac{-b + \sqrt{b^2 - 4a(c-y)}}{2a}\right)\frac{1}{\sqrt{b^2-4a(c-y)}} +$$

$$f_X\left(\frac{-b - \sqrt{b^2 - 4(c-y)}}{2a}\right)\frac{1}{\sqrt{b^2-4a(c-y)}} \quad .$$

Ejemplo:

Sea $X \sim N(0,1)$ ¿Cuál sería la distribución de $g(X)=X^2$?. En este caso a=1, b=0, c=0, aplicando la proposición 4, resulta que

$$f_Y(y)=\frac{1}{2\sqrt{y}}\left\{f_X(\sqrt{y}) + f_X(-\sqrt{y})\right\} =$$

$$= \frac{1}{2\sqrt{y}}\left\{\frac{1}{\sqrt{2\Pi}}e^{-y/2} + \frac{1}{\sqrt{2\Pi}}e^{-y/2}\right\} =$$

$$= \frac{2}{2\sqrt{2\pi y}}e^{-y/2}$$

por lo tanto

$$f_Y(y) = \begin{cases} \dfrac{1}{\sqrt{2\pi y}}e^{-y/2} & y>0 \\ 0 & \text{en otro caso} \end{cases}$$

la cual es conocida como un $X^2(1)$ (Chi-cuadrada con un grado de libertad).

Ahora bien para el caso en que la variable aleatoria sea discreta se tiene lo siguiente.

Proposición 5.-

Si X es una variable aleatoria discreta con función de densidad $f_x(x)$ y $Y=g(X)$ entonces la función de densidad de Y esta dada como $f_Y(y)=P(Y=y)=P(X=g^{-1}y))$.

Demostración:

Sea $T_y=(Y:Y=y)$ y $f_Y(y)$ la función de densidad de Y, entonces $f_Y(y)=P(Y=y)=P(Y\epsilon T)$ pero como $Y=g(X)$ resulta que $f_Y(y)=P(Y\epsilon T)=P(X=g^{-1}(y))$. ∎

Ejemplo:

Supóngase que se tiene una variable aleatoria discreta X que se distribuye binomialmente con parámetros $(4,1/2)$; sea $Y=(X-np)^2$ entonces ¿Cuál es la probabilidad de Y si X toma los valores 0,1,2,3,4?.

Como $Y=(X-np)^2 \rightarrow X=np \pm \sqrt{Y}$ aplicando la proposición 5 implica que $f_Y(y) = f_X(np \pm \sqrt{Y})$

$$= \binom{n}{np+\sqrt{y}} p^{np+\sqrt{y}} (1-p)^{n-(np+\sqrt{y})} +$$

$$\binom{n}{np-\sqrt{y}} p^{np-\sqrt{y}} (1-p)^{n-(np-\sqrt{y})}$$

entonces

$f_Y(0) = f_X(2) = \frac{6}{16}$

$f_Y(1) = f_X(1) + f_X(3) = \dfrac{4}{16} + \dfrac{4}{16} = \dfrac{8}{16}$

$f_Y(2) = 0$

$f_Y(3) = 0$

$f_Y(4) = f_X(0) + f_X(4) = \dfrac{1}{16} + \dfrac{1}{16} = \dfrac{2}{16}$.

Ahora bien si se analizan los resultados de las proposiciones, 1, 2, 4 se observa que si se incluye $F(\infty)$ y $F(-\infty)$, estos pueden expresarse como la suma de diferencias de los valores que toma la función de distribución acumulada, es decir; la proposición 1 puede expresarse como

$$F_Y(y) = F_X(g^{-1}(y)) - F_X(-\infty)$$

la proposición 2, como

$$F_Y(y) = F_X(\infty) - F_X(g^{-1}(y))$$

y la proposición 4, como

$$F_Y(y) = \left(F_X(\infty) - F_X\left(\frac{-b + \sqrt{b^2 - 4a(c-y)}}{2a} \right) \right)$$
$$+ \left(F_X\left(\frac{-b - \sqrt{b^2 - 4a(c-y)}}{2a} \right) - F_X(-\infty) \right).$$

El siguiente Teorema formaliza estas observaciones.

<u>Teorema</u> <u>4.2.</u>

Si $Y=g(x)$ es cualquier función continua de una variable aleatoria X, entonces la función de distribución de Y, $F_Y(y)$ es:

$$F_Y(y) = \sum_j \left\{ F_x(g^{-1}_{j2}(y)) - F_x(g^{-1}_{j1}(y)) \right\}$$

en donde $g^{-1}_{ji}(y)$, $i=1,2$, $j=1,2$ son raices no necesariamente diferentes de la ecuación $x=g^{-1}(y)$ y se puede incluir a ∞ y $-\infty$ ó ambas si es que se requiere.

Demostración:

Sea $T_y=(Y:Y\le y)$ entonces $T^c_y=(Y:Y>y)$ y,

$$P(T^c_y)=1-F_Y(y)=P\left\{X\in g^{-1}(T^c_y)\right\}$$

ahora el problema es caracterizar el conjunto $\left[g^{-1}(T_y)\right]^c$. Como $g(x)$ es una función continua, entonces para todo número real x' que satisfaga $g(x')>y$ existe un intervalo $I(x')$ de máxima longitud tal que para toda $x\in I(x')$, $g(x)>y$ y $I(x')$ contiene a x'. Pero $I(x')$ de máxima longitud significa que para cualesquier otro intervalo I que contenga a x' y que satisfaga $g(x)>y$ para cada $x\in I$ entonces $I\subset I(x')$. Por lo tanto, se deduce que $\left[g^{-1}(T_y)\right]^c$ tiene una única representación es decir,

$$g^{-1}(T^c_y) = \bigcup_j I(x_j), \qquad I(x_{j1})\ I(x_{j2})=\emptyset \text{ para } j1\neq j2$$

los intervalos $I(x_j)$ son todos abiertos, debido a que si es infinito ó finito no contiene a sus puntos extremos; es decir si es finito, cualquier punto extremo $x^{\bullet}$ de cualquier intervalo, se tiene que $g(x^{\bullet})=y$. Por lo tanto, resulta que

$$P\left\{X\in g^{-1}(T^c_y)\right\} = P\left\{\bigcup_j (I(x_j))\right\} = \sum_j P(I(x_j))$$

utilizando la observación de que los puntos extremos de cada intervalo son $-\infty$, ∞ ó una solución de $g(x)=y$ entonces se puede escribir $I(x_j)=(g^{-1}_{j1}(y) < x_j < g^{-1}_{j2}(y))$, de donde

$$P\left\{X\epsilon\ g^{-1}(T^c_y)\right\} = \sum_j \left\{F_x(g^{-1}_{j1}(y))-F_x(g^{-1}_{j2}(y))\right\}$$

por lo tanto,

$$F_Y(y)=1-P\left\{X\epsilon\ g^{-1}(T_y)^c\right\}=1-\sum_j\left\{F_X(g^{-1}_{j1}(y))\ -F_X(g^{-1}_{j2}(y))\right\}$$

Como $F(\infty)=1$, el "1" puede ser incluido dentro de la suma, por lo que resulta

$$F_Y(y)=\sum_j\left\{F_x(g^{-1}_{j2}(y))-F_x(g^{-1}_{j1}(y))\right\}. \qquad \blacksquare$$

<u>Teorema 4.3.</u>

Si X tiene función de densidad de probabilidad $f_x(x)$ y $g^{-1}_j(y)$ tiene derivada continua para toda j, distinta de cero en el rango de $g^{-1}_j(y)$, entonces

$$f_Y(y)=\sum f_X(g^{-1}_j(y))\left|g^{-1}_j(y)'\right|$$

donde la suma corre sobre todas las soluciones de $g(x)=y$ y se supone que solo hay a lo más un número determinado de inversas.

Demostración:

$$f_Y(y)=F_Y{}'(y)=\frac{d}{dy}\left\{\sum_j\left[F_X(g_{j2}(y))-F_X(g^{-1}_{j1}(y))\right]\right\} =$$

$$= \sum_j\left\{f_X(g_{j2}(y))(g^{-}_{j2}(y))'-f_X(g_{j1}(y))(g^{-}_{j1}(y))'\right\}. \qquad (1)$$

supóngase que $g_{j2}(y)\neq g_{j1}(y)$ para toda j, como $g_{j2}(y)$ es el punto extremo más pequeño de algún intervalo que satisface $g(x)>y$ es claro que $(g_{j2}^{-1}(y))'\geq 0$.

De manera similar se observa que $g_{j1}(y)\leq 0$, por lo tanto para toda j en la suma (1) son no negativos, por lo que se puede reemplazar (1) por:

$$f_Y(y)=\sum_j f_X(g_j(y))\,|(g_j^{-1}(y))'| \tag{2}$$

donde la suma corre sobre todas la soluciones de $g(x)=y$.

Si para algún j $g_{j2}^{-1}(y)=g_{j1}^{-1}(y)$ entonces se tiene un punto crítico y ese término se anula de la suma. $\blacksquare$

La fórmula (2) es conocida en Cálculo como la Fórmula de Cambio de Variable para integrales definidas.

Ejemplo:

Sea X variable aleatoria que se distribuye normalmente con parámetros $(0,\sigma^2)$. Si $Y=|X|$ ¿Cuál es su función de densidad?.

$$Y=g(X)=X=\begin{cases} x & \text{si } x\geq 0 \\ -x & \text{si } x<0 \end{cases}$$

entonces $g_1^{-1}(Y)=y$ si $x\geq 0$, y $g_2^{-1}(Y)=-y$ si $x<0$, utilizando (2) del teorema anterior resulta

$$f_Y(y)=\begin{cases} f_X(y)+f_X(-y) & y\geq 0 \\ 0 & y<0 \end{cases}$$

como $X\sim N(0,\sigma^2)$ entonces .

$$f_Y(y) = \frac{1}{\sigma\sqrt{2\pi}} e^{-y^2/2\sigma^2} + \frac{1}{\sigma\sqrt{2\pi}} e^{-(-y)^2/2\sigma^2} =$$

$$= \frac{2}{\sigma\sqrt{2\pi}} e^{-y^2/2\sigma^2}$$

por lo tanto

$$f_Y(y) = \begin{cases} \dfrac{2}{\sigma\sqrt{2\pi}} e^{-y^2/2\sigma^2} & y \geq 0 \\ 0 & \text{en otro caso} \end{cases}$$

TRANSFORMACION DE VARIABLES EN N-DIMENSIONES.

El concepto de transformación de variables aleatorias se generaliza a vectores aleatorios, presentando gran utilidad, debido a que nos ayuda a encontrar tanto la función de distribución conjunta de vectores aleatorios, como la de la suma ó del cociente de variables aleatorias independientes e identicamente distribuidas. Primero se analiza cuando la transformación $g(.)$ es una función de $\mathbb{R}^n \to \mathbb{R}^n$; por ejemplo si $X=(X_1,X_2,X_3)$, entonces $Y=(X_1+X_2+X_3, X_1-X_2, X_1-X_3)$. Posteriormente, cuando la transformación es una función de $\mathbb{R}^n \to \mathbb{R}^m$ con $m<n$; en este caso se introducen nuevas variables aleatorias, $Y_{m+1}=g_{m+1},\ldots,Y_n=g_n$, con lo cual se hace que la transformación sea de $\mathbb{R}^n \to \mathbb{R}^n$, así se determina la función de distribución conjunta de $Y_1,\ldots,Y_n$, y finalmente se encuentra la función de densidad marginal de $Y_1,\ldots,Y_m$.

Teorema 4.4.

Si X es un vector aleatorio con función de distribución conjunta $F_X(x_1 \ldots x_n)$ y $Y=g(X)$ es una función cuyo rango es un subconjunto del espacio euclidiano n-dimensional, entonces

$$F_Y(y_1, \ldots y_n) = \int \cdots_{g^{-1}(T_y)} \cdots \int f_{x_1 \ldots x_n}(x_1, \ldots, x_n)\, dx_1, \ldots, dx_n$$

donde

$$T_y = (Y; Y_1 \leq y_1, \ldots, Y_n \leq y_n)$$

Demostración:

Sea $T_y = (Y : Y_1 \leq y_1, \ldots, Y_n \leq y_n)$ y $F_Y(y_1, \ldots, y_n)$ la función de distribución de Y, entonces

$$F_Y(y_1, \ldots, y_n) = P(Y_1 \leq y_1, \ldots, Y_n \leq y_n) = P(Y \in T_y) = P(X \in g^{-1}(T_y)) =$$

$$= \int \cdots_{g^{-1}(T_y)} \cdots \int f_{x_1 \ldots x_n}(x_1, \ldots, x_n)\, dx_1, \ldots, dx_n$$

donde $g^{-1}(T) = (x : g(x) \in T)$. ∎

Proposición 6.

Si X es un vector aleatorio discreto, entonces la función de densidad de $Y=g(X)$,

$$f_Y(y_1, \ldots, y_n) = P(Y=y) = P(X=g^{-1}(Y)).$$

Se supone que la transformación $g(X)$ es una función continua y que cualquier imagen de Y puede a lo más tener un número numerable de inversas. Por lo tanto se tiene una familia de transformaciones inversas es decir,

$$\left.\begin{array}{l} X_1=g_{1j}^{-1}(Y_1,\ldots,Y_n) \\ \vdots \\ X_n=g_{nj}^{-1}(Y_1,\ldots,Y_n) \end{array}\right\} \qquad j=1,2,\ldots$$

por lo tanto se puede reescribir el teorema 4.4 como:

Proposición 7.

Si $g(X)$ es una función continua, entonces

$$F_Y(y_1,\ldots,y_n)=\sum_j \int_{g_j^{-1}(T_y)} \int f_{X_1\ldots x_n}(x_1,\ldots,x_n)\,dx_1,\ldots,dx_n.$$

Proposición 8.

Si X es una vector aleatorio discreto y $g(X)$ es una función continua entonces,

$$P(Y_1=y_1,Y_2=y_2,\ldots,Y_n=y_n)=f_Y(y)=$$

$$=\sum_j P\left\{X_1=g_{1j}^{-1}(y_1,y_2,\ldots,y_n),X_2=g_{2j}^{-1}(y_1,y_2,\ldots,y_n),\ldots\right.$$

$$\left.X_n=g_{nj}^{-1}(y_1,y_2,\ldots,y_n)\right\}.$$

En el siguiente teorema se determina la forma de encontrar la función de densidad de la transformación.

<u>Teorema 4.5.</u>

Sea X un vector aleatorio continuo, con función de densidad $f_X(x)$ y supóngase que para $x_1=g_{1j}^{-1}(y_1,\ldots,y_n)$, $\delta g_{1j}^{1}/\delta y_k$ es continua para todo i,j,k y que cada uno de los jacobianos J_j de las transformaciones inversas

$$J_j = \begin{vmatrix} \frac{\delta g_{1j}^{-1}}{\delta y_1} & \cdots & \frac{\delta g_{1j}^{-1}}{\delta y_n} \\ \vdots & & \\ \frac{\delta g_{nj}^{-1}}{\delta y_1} & \cdots & \frac{\delta g_{nj}}{\delta y_n} \end{vmatrix} \neq 0$$

entonces para cada y en el rango de $g(x)$ hay un número numerable de inversas y

$$f_{Y_1 \ldots Y_n}(y_1, \ldots, y_n) = \sum_j f_{X_1, \ldots, X_n}\left(g_{1j}^{-1}(y_1, \ldots y_n),\ g_{2j}^{-1}(y_1, \ldots y_n),\ \ldots\right.$$

$$\left. g_{nj}^{-1}(y_1, \ldots y_n)\right) |J_j| .$$

Demostración.

Sea B un conjunto cualquiera; interesa conocer la $P(Y \epsilon B)$, entonces se tiene que

$$P(Y \epsilon B) = \int \cdot_B \cdot \int f_{Y_1 \ldots Y_n}(y_1, \ldots, y_n)\ dy_1 \ldots dy_n$$

por el teorema 4.4

$$= \int \cdot_{g^{-1}(B)} \cdot \cdot \int f_{X_1 \ldots X_n}(x_1, \ldots, x_n)\, dx_1 \ldots dx_n$$

$$= \sum_j \int \cdot_{g_j^{-1}(B)} \cdot \cdot \int f_{X_1 \ldots X_n}(x_1, \ldots, x_n)\, dx_1 \ldots dx_n$$

La transformacion para cada $g_j^{-1}(B)$ en el conjunto B es una transformación uno a uno. Aplicando el teorema de cambio de variable se tiene que

$$P(Y\in B)=\sum_j \int\cdot\cdot_B\cdot\int f_{X_1,\ldots X_n}\left(g_{1j}^{-1}(y_1,\ldots y_n),\ g_{2j}^{-1}(y_1,\ldots y_n),\ \ldots\ g_{nj}^{-1}(y_1,\ldots y_n)\right)|J_j|\ dy_1\ldots dy_n$$

si $B=T_y$ donde $T_y=(y\in\mathbb{R}^n|Y_n\leq y_n)$ resulta que

$$P(Y\leq y)=\sum_j \int\int f_{X_1,\ldots X_n}\left(g_{1j}^{-1}(y_1,\ldots y_n),\ g_{2j}^{-1}(y_1,\ldots y_n),\ \ldots\ g_{nj}^{-1}(y_1,\ldots y_n)\right)|J_j|\ dy_1\ldots dy_n$$

por lo tanto

$$f_{Y_1\ldots Y_n}(y_1,\ldots,y_n)=\sum_j f_{X_1,\ldots X_n}\left(g_{1j}^{-1}(y_1,\ldots y_n),\ g_{2j}^{-1}(y_1,\ldots y_n),\ \ldots\ g_{nj}^{-1}(y_1,\ldots y_n)\right)|J_j|.\ \blacksquare$$

Ejemplos:

i) Supóngase que se tienen X_1 y X_2, variables aleatorias distribuidas normalmente con parámetros (0,1). Sea $Y_1=X_1+X_2$ y $Y_2=X_1/X_2$ entonces ¿Cuál es la función de densidad conjunta de (Y_1,Y_2)?.

Primero se determinan las funciones inversas de la transformación,

$$x_1=g_1^{-1}(y_1,y_2)=\frac{y_1y_2}{1+y_2} \quad y \quad x_2=g_2^{-1}(y_1,y_2)=\frac{y_1}{1+y_2}$$

$$J_j=\begin{vmatrix}\dfrac{y_2}{1+y_2} & \dfrac{y_1}{(1+y_2)^2}\\ \dfrac{1}{1+y_2} & \dfrac{y_1}{(1+y_2)^2}\end{vmatrix}=-\frac{y_1(y_2+1)}{(1+y_2)^3}=-\frac{y_1}{(1+y_2)^2}$$

sustituyendo resulta

$$f_{Y_1Y_2}(y_1,y_2)=\frac{y_1}{(1+y_2)^2}\ \frac{1}{2\pi}\exp\left\{-\frac{1}{2}\left[\frac{(y_1y_2)^2}{(1+y_2)^2}+\frac{y_1^2}{(1+y_2)^2}\right]\right\}$$

Ahora, si se encuentra la función de densidad marginal con respecto a Y_2, se obtiene la función de densidad Cauchy esto es,

$$f_{Y_2}(y_2)=\int_{-\infty}^{\infty}f_{Y_1,Y_2}(y_1,y_2)\,dy_1$$

$$=\frac{1}{(1+y_2)^2}\ \frac{1}{2\Pi}\int_{-\infty}^{\infty}|y_1|\exp\left\{-\frac{1}{2}\ \frac{(1+y_2^2)\ \ y_1^2}{(1+y_2)^2}\right\}dy_1$$

haciendo un cambio de variable

$$u=\frac{1}{2}\ \frac{(1+y_2^2)\ \ y_1^2}{(1+y_2)^2}\quad\text{entonces } du=\frac{(1+y_2^2)}{(1+y_2)^2}y_1dy_1$$

por lo tanto

$$f_{Y_2}(y_2)=\frac{1}{(1+y_2)^2}\ \frac{1}{2\Pi}\ \frac{(1+y_2)^2}{(1+y_2^2)}(2)\int_0^{\infty}e^{-u}du=\frac{1}{\Pi}\ \frac{1}{1+y_2^2}$$

entonces el cociente de dos variables independientes distribuidas normalmente tiene una función de densidad Cauchy.

ii) Sean $X_1,\ldots,X_k,X_{k+1}$ variables aleatorias independientes tales que cada $X_i\sim\Gamma(\alpha_i,1)$ y sea $Y_1=\dfrac{X_1}{X_1+\ldots+X_n}$, $Y_2=\dfrac{X_2}{X_1+\ldots+X_n}$, ...,

$Y_k=\dfrac{X_k}{X_1+\ldots+X_n}$, $Y_{k+1}=X_1+\ldots+X_{k+1}$; ¿Cuál es $f_{Y_1\ldots Y_k}(y_1,\ldots,y_k)$?.

Se conoce que

$$f_{X_1 \ldots X_{k+1}}(x_1, \ldots, x_{k+1}) = \begin{cases} \dfrac{x_1^{\alpha_1-1} x_2^{\alpha_2-1} \cdots x_{k+1}^{\alpha_{k+1}-1}}{\Gamma(\alpha_1) \cdots \Gamma(\alpha_{k+1})} e^{-(x_1+\ldots+x_{k+1})} & x_i \geq 0 \\ 0 & \text{e.o.c.} \end{cases}$$

pero se quiere encontrar $g_{ij}^{-1}(y_1, \ldots, y_{k+1}) = x_i$ entonces

$$g_1^{-1}(y_1, y_2, \ldots, y_{k+1}) = y_1 y_{k+1}$$
$$g_2^{-1}(y_1, y_2, \ldots, y_{k+1}) = y_2 y_{k+1}$$
$$\vdots$$
$$g_k^{-1}(y_1, y_2, \ldots, y_{k+1}) = y_k y_{k+1}$$
$$g_{k+1}^{-1}(y_1, y_2, \ldots, y_{k+1}) = y_{k+1}(1 - y_1 - y_2 - \ldots - y_k)$$

la transformación solo tiene una inversa. Ahora se calcula el jacobiano

$$J = \begin{vmatrix} y_{k+1} & 0 & \ldots & 0 & y_1 \\ 0 & y_{k+1} & \ldots & 0 & y_2 \\ \vdots & & & & \\ -y_{k+1} & -y_{k+1} & \ldots & -y_{k+1} & (1 - y_1 - \ldots - y_k) \end{vmatrix} = (y_{k+1})^k$$

entonces

$$f_{Y_1 \ldots Y_{k+1}}(y_1, \ldots, y_{k+1}) = f_{x_1 \ldots x_{k+1}}\left(g_1^{-1}, g_2^{-1}, \ldots, g_k^{-1}, g_{k+1}^{-1}\right)|J| =$$

$$= \frac{(y_1\, y_{k+1})^{\alpha_1-1} \cdots (y_k\, y_{k+1})^{\alpha_k-1} \;\; [y_{k+1}(1-y_1, \ldots y_k)]^{\alpha_{k+1}-1}}{\Gamma(\alpha 1)\Gamma(\alpha 2) \ldots \Gamma(\alpha k)\Gamma(\alpha k+1)} e^{-y_{k+1}}$$

$$y_{k+1}^k$$

para $y_i > 0$ $i = 1, \ldots, k$, si $y_1 + \ldots + y_k < 1$, $y_{k+1} > 0$ entonces

$f_{Y_1\ldots Y_{k+1}}(y_1,\ldots,y_{k+1}) =$

$$=\frac{(y_1)^{\alpha_1-1}\ldots(y_k)^{\alpha_k-1}(1-y_1,\ldots y_k)^{\alpha_{k+1}-1}}{\Gamma(1)\Gamma(\alpha_2)\ldots\Gamma(\alpha_{k+1})}(y_{k+1})^{(\alpha_1-1)\ldots(\alpha_{k+1}-1)}\ e^{-y_{k+1}}.$$

Ahora bien, lo que se desea es definir $f_{Y_1\ldots Y_k}(y_1,\ldots,y_k)$, por lo que se integra $f_{Y_1\ldots Y_{k+1}}(y_1,\ldots,y_{k+1})$ sobre Y_{k+1} es decir,

$$f_{Y_1\ldots Y_k}(y_1,\ldots,y_k)=\int_{Y_{k+1}} f_{Y_1\ldots Y_{k+1}}(y_1,\ldots,y_{k+1})\,dy_{k+1}$$

por lo que finalmente

$$f_{Y_1\ldots Y_k}(y_1,\ldots,y_k)=\begin{cases}\dfrac{\Gamma(\alpha_1+\ldots+\alpha_{k+1})}{\Gamma(\alpha_1)\ldots\Gamma(\alpha_{k+1})}y_1^{\alpha_1-1}\ldots y_k^{\alpha_k-1}(1-y_1\ldots y_k)^{\alpha_{k+1}-1} & \text{si } y_i>0,\ y_1+\ldots+y_k<1,\\ 0 & \text{en otro caso}\end{cases}$$

a esta distribución se le conoce como Dirichlet.

Caso m<n.

En algunas ocasiones puede ocurrir que sea de interes conocer el valor de una suma, de un cociente de variables aleatorias ó simplemente se quiera conocer el valor de una función de $\mathbb{R}^n \rightarrow \mathbb{R}^m$. Para ello se enuncia el siguiente teorema.

<u>Teorema 4.6.</u>

Sea $X=(X_1,X_2,\ldots,X_n)$ un vector aleatorio absolutamente continuo con función de densidad continua $f_{X_1\ldots X_n}(x_1,\ldots,x_n)$. Si $Y=(Y_1,Y_2,\ldots,Y_m)=g(X)=\left(g_1(X_1,\ldots,X_n),\ldots,g_m(X_1,\ldots X_n)\right)$ con m<n, se define una transformación auxiliar $Y^*=(Y_1,\ldots,Y_m,\ldots Y_n)$ donde cada

$Y_1=g_1(X_1,...,X_n)$, entonces para cada $x_1=g_{1j}^{-1}(y_1,...,y_n)$ si $\delta g_{ij}^{-1}/\delta y_k$ es continua para todo i,j,k y cada uno de los jacobianos J_j son distintos de cero en el rango de $g(x)$ se tiene que

$$f_{Y_1,...Y_n}(y_1,...y_n)=$$
$$=\sum\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} f_{X_1,...X_n}\left(g_{1j}^{-1}(y_1,...,y_n),...,\right.$$
$$\left.g_{nj}^{-1}(y_1,...,y_n)\right)|J_j|dy_{m+1}dy_{m+2}...dy_n.$$

Casos Particulares.

I) Distribución de la suma de variables aleatorias independientes.

Sean X_1, X_2 dos variables aleatorias con función de densidad $f_{X_1X_2}(x_1,x_2)$, ¿Cuál sería la función de densidad de $Y=X_1+X_2$?. En este caso la transformación es de $\mathbb{R}^2\rightarrow\mathbb{R}$, para poder aplicar el teorema 4.6 se necesita definir una transformación auxiliar por lo que se hace lo siguiente:

$$Y=X_1+X_2=g_1(X_1,X_2)$$
$$Z=\ X_2\ =g_2(X_1,X_2)$$

entonces

$$g_2^{-1}(y,z)=z \text{ y } g_1^{-1}(y,z)=y-z$$

por lo que

$$J_j=\begin{vmatrix}\dfrac{\delta g_1^{-1}}{\delta y} & \dfrac{\delta g_1^{-1}}{\delta z}\\[2ex] \dfrac{\delta g_2^{-1}}{\delta y} & \dfrac{\delta g_2}{\delta z}\end{vmatrix}=\begin{vmatrix}1 & -1\\[2ex] 0 & 1\end{vmatrix}=|\ 1\ |=1$$

sustituyendo valores resulta

$$f_{YZ}(y,z)=f_{X1X2}(g_1^{-1}(y,z),g_2^{-1}(y,z))\,|J|=f_{X1X2}(y-z,z)$$

por lo tanto

$$f_Y(y)=\int_{-\infty}^{\infty} f_{X1}(y-z)f_{X2}(z)\,dz. \qquad (3)$$

Ejemplo:

Sean X_1, X_2 variables aleatorias independientes tales que $X_1 \sim N(\mu_1,\sigma_1^2)$ y $X_2 \sim N(\mu_2,\sigma_2^2)$ entonces ¿Cuál es la distribución de $Y=X_1+X_2$?.

Aplicando (3) resulta que

$$f_Y(y)=\int_{-\infty}^{\infty} f_{X1}(y-x_2)f_{X2}(x_2)\,dx_2$$

$$=\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\,\sigma_1} e\,xp^{-(y-x_2-\mu_1)^2/2\sigma_1^2}\ \frac{1}{\sqrt{2\pi}\,\sigma_2}\exp^{-(x_2-\mu_2)/2\sigma_1^2}\,dx_2$$

$$=\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\,\sigma_1\sigma_2}\exp\frac{1}{2}\left[-(y-x_2-\mu_1)^2/2\sigma_1^2-(x_2-\mu_2)^2/2\sigma_2^2\right]dx_2$$

considérese por el momento solo el exponencial,

$$-\frac{(y-x_2-\mu_1)^2}{2\sigma_1^2}-\frac{(x_2-\mu_2)^2}{2\sigma_2^2}$$

si se factoriza resulta que

$$=-\frac{(\sigma_1^2+\sigma_2^2)}{2\sigma_1^2\sigma_2^2}\left[x_2^2-2x_2\left(\frac{\sigma_1^2\mu_2+\sigma_2^2(y-\mu_1)}{\sigma_1^2+\sigma_2^2}\right)\right]$$

$$-\frac{1}{2}\left[\frac{(y-\mu_1)^2}{\sigma_1^2}+\frac{\mu_2^2}{\sigma_2^2}\right],$$

completando el cuadrado en x_2 queda

$$= -\frac{(\sigma_1^2+\sigma_2^2)}{2\sigma_1^2\sigma_2^2}\left[\left(x_2-\frac{\sigma_1^2\mu_2+\sigma_2^2(y-\mu_1)}{\sigma_1^2+\sigma_2^2}\right)^2\right]$$

$$+\frac{(\sigma_1^2+\sigma_2^2)}{2\sigma_1^2\sigma_2^2}\left(\frac{\sigma_1^2\mu_2+\sigma_2^2(y-\mu_1)}{\sigma_1^2+\sigma_2^2}\right)^2-\frac{1}{2}\left[\frac{(y-\mu_1)^2}{\sigma_1^2}+\frac{\mu_2^2}{\sigma_2^2}\right].$$

por lo tanto

$$f_Y(y)=\frac{\exp\left\{-\frac{1}{2}\left[\frac{(y_1-\mu_1)^2}{\sigma_1^2}+\frac{\mu_2^2}{\sigma_2^2}\right]+\frac{(\sigma_1^2\mu_2+\sigma_2^2(y-\mu_1))^2}{(2\sigma_1^2\sigma_2^2)(\sigma_1^2+\sigma_2^2)}\right\}}{2\pi\sigma_1\sigma_2}$$

$$\int_{-\infty}^{\infty}\exp\left\{-\frac{(\sigma_1^2+\sigma_2^2)}{2\sigma_1^2\sigma_2^2}\left(x_2-\frac{\sigma_1^2\mu_2+\sigma_2^2(y-\mu_1)}{\sigma_1^2+\sigma_2^2}\right)^2\right\}dx_2$$

la integral es una normal con media $\frac{\sigma_1^2\mu_2+\sigma_2^2(y-\mu_1)^2}{\sigma_1^2+\sigma_2^2}$ y varianza

$\frac{\sigma_1^2\ \sigma_2^2}{(\sigma_1^2+\sigma_2^2)}$ con la constante$(1/\sqrt{2\pi})(\sqrt{(\sigma_1^2+\sigma_2^2)/(\sigma_1^2\sigma_2^2)}$,

por lo tanto

$$f_Y(y)=\frac{\exp\left\{-\frac{1}{2}\left[\frac{(y_1-\mu_1)^2}{\sigma_1^2}+\frac{\mu_2^2}{\sigma_2^2}\right]+\frac{(\sigma_1^2\mu_2+\sigma_2^2(y-\mu_1))^2}{(2\sigma_1^2\sigma_2^2)(\sigma_1^2+\sigma_2^2)}\right\}}{\sqrt{2\pi(\sigma_1^2+\sigma_2^2)}}$$

si se acomodan nuevamente los términos del exponencial resulta que

$$f_Y(y)=\frac{\exp\left\{-\frac{1}{2}\left[\frac{(y_1-\mu_1-\mu_2)^2}{(\sigma_1{}^2+\sigma_2{}^2)}\right]\right\}}{\sqrt{2\pi(\sigma_1{}^2+\sigma_2{}^2)}},$$

que es una normal con media $\mu_1+\mu_2$ y varianza $\sigma_1^2+\sigma_2^2$.

II) Distribución del cociente de variables aleatorias independientes.

En el caso del cociente de variables aleatorias $Y=X_1/X_2$ ¿Cuál sería la función de densidad de $f_Y(y)$?.

Al igual que en el caso de la suma se completa la transformación para que sea de $\mathbb{R}^2\to\mathbb{R}^2$, es decir,

$Y=X_2/X_1$ con $(X_1\neq 0)$
$Z=X_1$

entonces

$$g_2^{-1}(y,z)=z \text{ y } g_1^{-1}(y,z)=yz$$

por lo que

$$J_J=\begin{vmatrix}\dfrac{\delta g_1^{-1}}{\delta y} & \dfrac{\delta g_1^{-1}}{\delta z}\\ \dfrac{\delta g_2^{-1}}{\delta y} & \dfrac{\delta g_2}{\delta z}\end{vmatrix}=\begin{vmatrix}0 & 1\\ z & y\end{vmatrix}=|-z| = z$$

sustituyendo valores resulta

$$f_Y(y)=\int_{-\infty}^{\infty} f_{X_1X_2}(z,yz)\,|z|dz$$

que es la fórmula para el cociente de dos variables aleatorias independientes.

Distribución F.

Sean X_1, X_2 variables aleatorias tales que $X_1 \sim X^2_{(n_1)}$ y $X_2 \sim X^2_{(n_2)}$ entonces ¿Cuál será la función de distribución de

$$Y = \frac{X_1/n_1}{X_2/n_2} = \frac{n_2\ X_1}{n_1\ X_2}\ ?.$$

Sea $W = \dfrac{X_1}{X_2}$ entonces primero se calculará la función de densidad de W y luego se calcula la función de densidad de $Y = W\dfrac{n_2}{n_1}$

$$f_W(w) = \int_0^\infty f_{X_1}(wx_2) f_{X_2}(x_2)\,|x_2|\,dx_2$$

$$= \int_0^\infty \left\{ \left(\frac{1}{2}\right)^{n_1/2} \frac{(wx_2)^{n_1/2 - 1} \exp^{-wx_2/2}}{\Gamma(^{n_1}/2)} \right.$$

$$\left. \left(\frac{1}{2}\right)^{n_2/2} \frac{(x_2)^{n_2/2 - 1} \exp^{-x_2/2}}{\Gamma(^{n_1}/2)} \right\} |x_2|\ dx_2$$

entonces

$f_W(w) =$

$$\int_0^\infty \left(\frac{1}{2}\right)^{n_1+n_2/2} \frac{w^{n_1/2 - 1} x_2^{\ n_1+n_2/2 - 1} \exp^{-x_2/2(1+w)}}{\Gamma(^{n_1}/2)\Gamma(^{n_2}/2)}\,dx_2$$

$$= \frac{1}{2^{n_1+n_2/2}} \quad \frac{w^{n_1/2 - 1}}{\Gamma(^{n_1}/2)\Gamma(^{n_2}/2)} \int_0^\infty x_2^{\frac{n_1+n_2}{2} - 1} \exp^{-x_2(1+w/2)}\ dx_2$$

$$= \frac{1}{2^{n_1+n_2/2}\,\Gamma(^{n_1}/_2)\Gamma(^{n_2}/_2)} \frac{w^{n_1/2-1}\,\Gamma(n_1+n_1/2)}{\left(\frac{1+w}{2}\right)^{n_1+n_2/2}}$$

por lo que

$$f_W(w)=\begin{cases} \dfrac{w^{\frac{n_1}{2}-1}\,\Gamma\left(\frac{n_1+n_2}{2}\right)}{2^{n_1+n_2/2}\Gamma(n_1/2)\Gamma(n_2/2)(1+w/2)^{n_1+n_2/2}} & w>0 \\ 0 & \text{e.o.c.} \end{cases}$$

sea $Y=\frac{n_2}{n_1}W$, $g^{-1}(y)=\frac{n_1}{n_2}y$, $\left(g^{-1}(y)\right)'=\frac{n_1}{n_2}$ entonces

$$f_Y(y)=f_W(g^{-1}(y))\,|(g^{-1}(y))'|$$

sustituyendo resulta

$$f_Y(y)=f_W\left(\frac{n_1}{n_2}y\right)\frac{n_1}{n_2}$$

$$= \frac{\left(\frac{n_1}{n_2}\right)^{\frac{n_1}{2}-1}\left(y\right)^{\frac{n_1}{2}-1}\Gamma\left(\frac{n_1+n_2}{2}\right)}{2^{n_1+n_2/2}\,\Gamma(n_1/2)\Gamma(n_2/2)(1+(n_1/n_2)y/2)^{n_1+n_2/2}}\;\frac{n_1}{n_2}$$

por lo tanto

$$f_Y(y)=\begin{cases} \dfrac{n_1^{\frac{n_1}{2}}\,\Gamma\left(\frac{n_1+n_2}{2}\right)\left(y\right)^{\frac{n_1}{2}-1}}{n_2^{\frac{n_1}{2}}\,\Gamma(n_1/2)\Gamma(n_2/2)(1+(n_1/n_2)y)^{\frac{n_1+n_2}{2}}} & y>0 \\ 0 & \text{e.o.c.} \end{cases}$$

a esta función de distribución de probabilidad se le conoce como la distribución F con n_1 y n_2 grados de libertad. Se denota como $Y\sim F(n_1,n_2)$.

Distribución T.

Sean X_1, X_2 dos variables aleatorias tales que $X_1 \sim N(0,1)$ y $pX_2 \sim X^2(p)$, ¿Cuál es la distribución de la variable $Y = X_1/\sqrt{X_2}$?.

Sea $W = pX_2$ entonces la función de densidad de W esta definida por:

$$g_W(w) = \frac{e^{-w/2} w^{(p/2)-1}}{2^{p/2}\,\Gamma(p/2)}, \qquad w>0$$

se sabe que $f_Y(y) = \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right)$ si $g(X) = aX+b$ entonces si se sustituye resulta que

$$f_{X_2}(x) = \frac{p e^{-px/2} (px)^{(p/2)-1}}{2^{p/2}\,\Gamma(p/2)}, \qquad x>0$$

Sea $U = \sqrt{X_2}$, se puede aplicar la proposición 1, debido a que se restringe a que $x>0$, entonces para $u>0$;

$$f_U(x) = \frac{p e^{-pu^2/2} (pu^2)^{(p/2)-1}}{2^{p/2}\,\Gamma(p/2)} (2u)$$

$$= \frac{p^{p/2} e^{pu^2/2} u^{p-1}}{2^{(p/2)-1}\,\Gamma(p/2)}$$

Ahora si $Y_1 = (X/U)$ y $Y_2 = U$, y se aplica la fórmula para el cociente resulta

$$f_Y(y_1) = \int_0^{\infty} \frac{y_2}{\sqrt{2\Pi}} \exp\left[-\frac{(y_1 y_2)^2}{2}\right] \frac{p^{p/2} \exp\left(-\frac{p y_2^2}{2}\right) y_2^{p-1}}{2^{(p/2)-1}\,\Gamma(p/2)} \, dy_2$$

$$=\int_0^{\infty} \frac{p^{p/2} \exp\left(-\frac{y_2^2}{2}(y_1^2+p)\right) y_2^p}{2^{(p/2)-1}\Gamma(p/2)\sqrt{2\Pi}}\,dy_2$$

para evualarla se hace $v=y_2^2$ entonces $dy_2=\frac{1}{2}\ v^{-1/2}dv$, sustituyendo en la integral resulta

$$=\int_0^{\infty} \frac{p^{p/2} \exp -(v/2)(y_1^2+p)\, v^{p/2}}{2^{(p/2)-1}\sqrt{2\Pi}\ \Gamma(p/2)}\,(1/2)v^{-1/2}dv$$

evaluando queda

$$f_Y(y_1)=\frac{p^{p/2}\ \Gamma[(p+1)/2]2^{(p+1)/2}}{2^{p/2}\sqrt{2\Pi}\ \Gamma(p/2)(y_1^2+p)^{(p+1)/2}}$$

$$f_Y(y_1)=\frac{\Gamma[(p+1)/2][1+(y_1^2/p)]^{-(p+1)/2}}{\Gamma(p/2)\sqrt{\Pi p}}$$

a $f_{Y_1}(y)$ se le conoce como la distribución "t" de Student con p grados de libertad.

Como se ha podido observar las transformaciones de variables aleatorias son muy utiles, sin embargo, su desarrollo es muy largo y tedioso. En el capitulo VI se analiza otro método que también puede ser utilizado para encontrar la función de distribución de variables aleatorias definidas en término de otras en las que se conoce como se distribuyen.

CARACTERISTICAS NUMERICAS DE VARIABLES Y VECTORES ALEATORIOS

En los capítulos anteriores se ha visto como las funciones de distribución determinan la probabilidad con que las variables aleatorias toman diferentes valores; sin embargo en algunas ocasiones solo se quiere conocer ciertas características numéricas de las funciones de distribución como son la Esperanza Matemática, Varianza, Coeficiente de Correlación, Función Generatriz de Momentos.

En este capítulo se analizan las características numéricas, primero para variables aleatorias y posteriormente se presenta la generalización de los conceptos a vectores aleatorios.

I. Variables Aleatorias.

Para introducir el concepto de esperanza matemática, considérese el siguiente ejemplo. Supóngase que se tiene un dado equiprobable, tirarlo cuesta 4 pesos, pero a cambio se recibe el número que cae en el dado. ¿Cuál es la ganancia promedio de tirar el dado?. Cada lado tiene probabilidad 1/6 entonces

$$1*1/6+2*1/6+3*1/6+4*1/6+5*1/6+6*1/6=3.5$$

este número indica que la ganancia promedio es de 3.5, se le conoce como la esperanza matemática. La siguiente definición formaliza este concepto,

Definición 5.1.

Sea $(\Omega\ \mathcal{A}.P)$ un espacio de probabilidad, X variable aleatoria definida sobre ese mismo espacio entonces se llama *Media o Esperanza Matemática* de X al número $E(X)$ (también denotada como μ_X),

$$E(X) = \int_{-\infty}^{\infty} x \, f_X(x)\,dx$$

si X es una variable continua, y la integral

$$\int_{-\infty}^{\infty} |x| \;\; f_x(x)\,dx$$

existe; y

$$E(X) = \sum x \, f_x(x)$$

si X es una variable discreta, y la serie converge absolutamente.

Como se puede observar $E(X)$ es un "promedio" de los valores que la variable aleatoria toma, cada valor está influenciado por la probabilidad de que la variable aleatoria tome ese valor, y por supuesto, aquellos valores que son más probables reciben mayor peso. La esperanza matemática es el centro de gravedad de la función de densidad de la variable aleatoria X, por lo tanto, es una medida que indica donde se centran los valores que toma la variable aleatoria.

Existen casos en los que las funciones de densidad no tiene esperanza matemática finita como es el caso de la función Cauchy,

$$f(x) = \frac{1}{\Pi(1 + x^2)} \qquad -\infty<x<\infty$$

ya que

$$\begin{aligned}
\int_{-\infty}^{\infty} x \frac{1}{\Pi(1 + x^2)}\,dx &= \frac{2}{\Pi} \int_0^{\infty} \frac{x}{1 + x^2}\,dx \\
&= \frac{2}{\Pi} \lim_{c\to\infty} \int_0^{c} \frac{x}{1 + x^2}\,dx \\
&= \frac{1}{\Pi} \lim_{c\to\infty} \log\,(1 + x^2)\,\Big|_0^{c} \\
&= \infty.
\end{aligned}$$

Ahora se enuncia un resultado que se utiliza en posteriores demostraciones.

Lema 5.1.

Para cualquier variable aleatoria Y,

$$E(Y)=\int_0^\infty P(Y>y)dy - \int_0^\infty P(Y<-y)dy.$$

Demostración.

Sea Y una variable aleatoria continua con función de densidad $f_Y(y)$ entonces

$$\int_0^\infty P(Y>y)dy = \int_0^\infty \int_y^\infty f_Y(x)dxdy$$

cambiando el orden de integración resulta

$$\int_0^\infty P(Y>y)dy = \int_0^\infty \left(\int_0^x dy \right) f_Y(x)dx$$

$$= \int_0^\infty x \, f_Y(x)dy,$$

igualmente

$$\int_0^\infty P(Y<-y)dy = \int_0^\infty \int_{-\infty}^{-y} f_Y(x)dxdy$$

$$\int_0^\infty P(Y<-y)dy = \int_{-\infty}^0 \left(\int_0^{-x} dy \right) f_Y(x)dx$$

$$=- \int_{-\infty}^0 x \, f_Y(x)dx$$

entonces sustituyendo

$$\int_0^{\infty} P(Y>y)dy - \int_0^{\infty} P(Y<-y)dy = \int_0^{\infty} x\, f_Y(x)dy + \int_{-\infty}^{0} x\, f_Y(x)dx$$

$$= \int_{-\infty}^{\infty} x\, f_Y(x)dx = E(Y). \quad \blacksquare$$

Es evidente que la esperanza matemática no puede de ningún modo ser característica suficiente de la variable aleatoria, ya que unicamente determina el valor medio cerca del cual se dispersan los valores posibles de dicha variable. Por lo que surge la *varianza* que completa la información aportada por la media, en el sentido de indicar cuál es la colocación de los valores observados alrededor de ella. Por ejemplo si se deseara invertir en la bolsa, seria conveniente no solo conocer el rendimiento promedio (esperanza matemática) de las acciones, sino también el grado de dispersión de los rendimientos alrededor de él, lo que involucra mayor ó menor grado de certidumbre ó riesgo. De tal forma que si el valor que tome la varianza es mayor que el de otra acción cuyo rendimiento no es tan atractivo, muy posiblemente una persona conservadora invertirá en la segunda acción.

<u>Definición 5.2.</u>

Sea X la variable aleatoria, $E(X)=\mu_x$, entonces la *varianza* de X denotada por σ_x^2 ó var[X] es

$$\mathrm{var}[X]=\sum (x-\mu_x)^2 f_X(x)$$

si X es una variable discreta; y

$$\mathrm{var}[X]=\int_{-\infty}^{\infty}(x-\mu_x)^2 f_X(x)dx$$

si X es una variable continua.

La varianza es una "medida de dispersión" ya que cuanto mayor sea el grado de concentración de los valores que toma la variable aleatoria alrededor de la media menor será el valor que tome la varianza. Las diferencias de cuadrados con mayor probabilidad tienen mayor peso.

Ejemplo:

Si se continua con el ejemplo anterior de arrojar un dado equiprobable entonces

$E(x)=3.5$

$$\mathrm{var}[X]=1/6\left[(1-3.5)^2+(2-3.5)^2+(3-3.5)^2+(4-3.5)^2+(5-3.5)^2+(6-3.5)^2\right]$$

$\mathrm{var}[X]=2.91.$

La varianza tiene la misma unidad que el cuadrado de la variable aleatoria. Sin embargo, seria conveniente que como "medida de dispersión" tuviera la misma unidad que la variable aleatoria en cuestión. Para obtener tal caracteristica se define a la desviación estándar, como la raiz cuadrada positiva de la varianza.

<u>Definición 5.3.</u>

Sea X una variable aleatoria, entonces la *desviación estándar* se define como $\sqrt{\sigma_x^2}$.

Al igual que la varianza, la desviación estándar es una medida de dispersión. Del ejemplo del dado que se ha venido analizando resulta que $\sigma_X=1.70$.

Ahora se calcula la esperanza y varianza de una X variable aleatoria que se distribuye normalmente;

$$f_x(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp - \left(\frac{(x-\mu)^2}{2\sigma^2} \right)$$

entonces

$$E(X) = \int_{-\infty}^{\infty} x \frac{1}{\sigma\sqrt{2\pi}} \exp - \left(\frac{(x-\mu)^2}{2\sigma^2} \right) dx$$

si se hace que $z=(x-\mu)/\sigma$ resulta que

$$E(X) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (\sigma z + \mu)\, e^{-z^2/2} dz$$

$$= \frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z e^{-z^2/2} dz + \frac{\mu}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-z^2/2} dz$$

como

$$\int_{-\infty}^{\infty} e^{-z^2/2} dz = \sqrt{2\pi} \quad \text{y} \quad \int_{-\infty}^{\infty} z e^{-z^2/2} dz = 0$$

entonces $E(X) = \mu$, para determinar la varianza se hace:

$$\text{var}(X) = \int_{-\infty}^{\infty} (x-\mu)^2 f_X(x)\, dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (x-\mu)^2 \exp \left(- \frac{(x-\mu)^2}{2\sigma^2} \right) dx$$

si se realiza un cambio de variable $z=(x-\mu)/\sigma$ resulta;

$$\text{var}(X) = \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z^2 e^{-z^2/2} dz$$

integrando por partes queda

$$\int_{-\infty}^{\infty} z^2 e^{-z^2/2} dz = -z e^{-z^2/2} \Big|_{-\infty}^{\infty} + \int_{-\infty}^{\infty} e^{-z^2/2} dz = \sqrt{2\pi}$$

por lo tanto $\text{var}(X)=\sigma^2$.

Proposición 5.1.

Si X es una variable aleatoria entonces

$$\mathrm{var}[X]=E[(X-E[X])^2]=E[X^2]-(E[X])^2.$$

La demostración se deduce de la definición. ∎

Proposición 5.2.

Si c es una constante entonces

$$\mathrm{Var}[cX]=c^2\mathrm{Var}[X].$$

Demostración:

$$\mathrm{var}[cX]=E[(cX-E[cX])^2]=E[(cX-cE[X])^2]$$
$$=c^2E[(X-E[X])^2].$$ ∎

VALOR ESPERADO DE UNA FUNCION DE UNA VARIABLE ALEATORIA.

En alguna ocasiones se desea encontrar la esperanza de una función de X, es decir la $E(X^2)$, o bien la $E(e^X)$, etc. ¿Como se determinaria?. Considérese el siguiente ejemplo, si X se distribuye normalmente con parámetros (0,1) ¿Cuál es la $E(X^2)$?; una forma de calcularlo sería como $g(X)=X^2$ es una variable aleatoria por si sola, entonces habría que calcular su función de distribución, una vez que se tuviera se podría calcular la $E(X^2)$. En el capítulo anterior se obtuvo que si $X\text{-}N(0,1)$ entonces la variable aleatoria $Y=g(X)=X^2$ se distribuye como una Chi-cuadrada con un grado de libertad entonces

$$E(Y)=\int_0^\infty y\ \frac{1}{\sqrt{2\pi}}\ y^{-1/2}\ e^{-1/2y}\,dy$$

integrando por partes

$$u=y^{1/2} \qquad dv=e^{-1/2y}dy$$

$$du=\frac{1}{2}\ y^{-1/2}dy \qquad v=-2e^{-1/2y}dy$$

resulta que

$$E(Y)=\frac{1}{\sqrt{2\pi}}\int_0^{\infty} y^{-1/2}\ e^{-1/2y}\ dy = \frac{1}{\sqrt{2\pi}}\ \frac{\Gamma(1/2)}{(1/2)^{1/2}}=1. \qquad \blacksquare$$

Por otro lado existe otro método que permite evaluar la esperanza de una función de una variable aleatoria sin tener primero que encontrar su función de distribución. El siguiente teorema establece la forma de hacerlo.

<u>Teorema</u> <u>5.1.</u> (Estadístico Inconciente)[1].

Sea X una variable aleatoria y $g:\mathbb{R}\rightarrow\mathbb{R}$ entonces

$$E[g(X)]=\int_{-\infty}^{\infty} g(X)f_X(x)dx$$

si X es una variable continua, y la integral

$$\int_{-\infty}^{\infty} |g(X)|\ f_X(x)dx$$

existe; y

$$E[g(X)]=\sum g(x)f_X(x)$$

si X es una variable discreta, y la serie converge absolutamente.

Demostración.

Sea X una variable aleatoria continua y $g:\mathbb{R}\rightarrow\mathbb{R}$ entonces aplicando el lema 5.1 se tiene que

---

[1] Sheldom Ross.- A First Course to the Theory of Probability.

$$E(g(X)) = \int_0^\infty P(g(X)>y)dy - \int_0^\infty P(g(X)<-y)dy$$

$$= \int_0^\infty \int_{x:g(x)>y} f(x)dxdy - \int_0^\infty \int_{x:g(x)<-y} f(x)dxdy$$

$$= \int_{x:g(x)>0} \int_0^{g(x)} dyf(x)dx - \int_{x:g(x)<0} \int_0^{-g(x)} dyf(x)dx$$

$$= \int_{x:g(x)>0} g(x)f(x)dx + \int_{x:g(x)<0} g(x)f(x)dx$$

$$= \int_{-\infty}^{\infty} g(x)f(x)dx.$$

Sea $Y=g(X)$, donde X una variable aleatoria discreta y $g:\mathbb{R}\to\mathbb{R}$ entonces

$$E(g(X)) = E(Y) = \sum_j y_j \; f_Y(y_j) = \sum g(x_j) \; f_{g(x)}(y_j)$$

$$= \sum_j g(x_j) \; P[g(X)=y_j]$$

$$= \sum_j g(x_j) \; P[X=g^{-1}(y_j)]$$

$$= \sum_j g(x_j) \; P[X=g^{-1}(g(y_j))]$$

$$= \sum_j g(x_j) \; P[X=x_j]$$

$$= \sum_j g(x_j) \; f_X(x_j). \qquad \blacksquare$$

Continuando con el ejemplo anterior si $X\text{-}N(0,1)$ entonces la esperanza de $g(X)=X^2$, aplicando el teorema anterior, queda

$$E(X^2) = \int_{-\infty}^{\infty} x^2 \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$$

integrando por partes

$$= \frac{1}{\sqrt{2\pi}}\left[-xe^{-x^2/2}\Big|_{-\infty}^{\infty} + \int_{-\infty}^{\infty} e^{-x^2/2}\,dx\right] = 1. \qquad \blacksquare$$

Como se puede observar, por ambos métodos se llega al mismo resultado. Ahora, se enuncian algunas propiedades de la esperanza matemática.

<u>Teorema 5.2.</u>

Algunas propiedades de la *Esperanza matemática* son:

i) $E[c]=c$ para c constante,
ii) $E[cg(X)]=cE[g(X)]$ para c constante,
iii) $E[c_1g_1(X)+c_2g_2(X)]=c_1E[g_1(X)]+c_2E[g_2(X)]$,
iv) $E[g_1(X)]\le E[g_2(X)]$ para $g_1(x)\le g_2(x)$.

Demostración:

i) X toma el valor de c con probabilidad 1, por definición $E(c)=c$.

ii) si X es continua entonces
$$E[cg(X)]=\int_{-\infty}^{\infty}cg(X)\,dF_X(x)=c\int_{-\infty}^{\infty}g(X)\,dF_X(x)=cE[g(X)].$$

iii) $E[c_1g_1(X)+c_2g_2(X)]=\int_{-\infty}^{\infty}[c_1g_1(x)+c_2g_2(x)]dF_X(x)=$

$$c_1\int_{-\infty}^{\infty}g_1(x)\,dF_X(x)+c_2\int_{-\infty}^{\infty}g_2(x)\,dF_X(x)=c_1E[g(X)]+c_2E[g(X)].$$

iv) $0\le E[g_2(X)-g_1(X)]=E[g_2(X)]-E[g_1(X)]$. $\blacksquare$

Ahora bien, si $g(x)=x$ entonces $E[g(X)]=E(X)$ que es la esperanza matemática de X. Si $g(x)=(x-\mu_x)^2$ entonces $E[g(x)]=E[(x-\mu_x)^2]=Var(X)$.

MOMENTOS.

Se ha definido la esperanza matemática de cualquier función real $g(X)$ de la variable aleatoria X. En particular, si $g(X)=x^k$ entonces $E[g(X)]=E[x^k]$ y se llama el momento no-central de orden k de la variable aleatoria X.

<u>Definición</u> <u>5.4.</u>

Si X es una variable aleatoria, entonces se define al k-ésimo momento no-central de X como $\mu_k'=E[X^k]$. Si X es una variable aleatoria continua y la integral existe

$$E[X^k]=\int_{-\infty}^{\infty}x^k f_X(x)\,dx$$

y si X es una variable discreta, y la serie converge absolutamente entonces

$$E[X^k]=\sum x\, f^k_X(x).$$

<u>Definición</u> <u>5.5.</u>

Si X es una variable aleatoria, entonces se define al k-ésimo momento central de X sobre a como $E[(X-a)^k]$. Si $a=\mu$,

$$E[(X-\mu)^k]=\mu_k$$

Los momentos de diferentes órdenes de una variable aleatoria pueden servir de características numéricas de la misma. Principalmente son utilizados, el primer momento no-central, el segundo momento central, puesto que el primero determina la esperanza de la variable aletoria en cuestión y el segundo la varianza de la misma.

Cuando la esperanza de la variable aleatoria es finita se cumple que el momento central de primer orden de la variable aleatoria X, es igual a cero.

El siguiente teorema establece la relación que existe entre los momentos no-centrales y centrales,

<u>Teorema 5.3.</u>

$$\mu_k = \sum_{j=0}^{k} \binom{k}{j} (-1)^j \mu^j \mu'_{k-j} \text{ para } k=0,1,\ldots$$

Demostración:

$$E((X-\mu)^k) = E\left\{ \sum_{j=0}^{k} \binom{k}{j} (-1)^j \mu^j X^{k-j} \right\}$$

$$E((X-\mu)^k) = \sum_{j=0}^{k} \binom{k}{j} (-1)^j \mu^j E\left\{ X^{k-j} \right\}$$

$$= \sum_{j=0}^{k} \binom{k}{j} (-1)^j \mu^j \mu'_{k-j}$$

en particular si:

$k=1$, $E((X-\mu)) = 0$

$k=2$, $E((X-\mu)^2) = \sigma^2 = \mu'_2 - 2\mu^2 + \mu^2 = \mu'_2 - \mu^2$ ∎

Ejemplo:

Sea X variable aleatoria que se distribuye con una Gamma ¿Cuál es su k-ésimo momento no central?.

$$\mu'_k = \int_0^{\infty} \frac{x^k e^{-\alpha x} \alpha x^{n-1}}{\Gamma(n)} dx = \int_0^{\infty} \frac{\alpha^n e^{-\alpha x} x^{n+k-1}}{\Gamma(n)} dx = \frac{\Gamma(n+k)}{\alpha^k \Gamma(n)}$$

por lo tanto la esperanza matemática es

$$\mu = \frac{n}{\alpha}$$

y la varianza es

$$\sigma^2 = \frac{n(n+1)}{\alpha^2} - \frac{n^2}{\alpha^2} = \frac{n}{\alpha^2}.$$

FUNCION GENERATRIZ DE MOMENTOS.

Sea X una variable aleatoria con función de densidad $f_X(\cdot)$. La función generatriz de momentos se define como el valor esperado de e^{tX}, siempre y cuando el valor esperado para cualquier t en el intervalo $-h<t<h$ con $h>0$, exista.

$$m(t)=E(e^{tX})= \int_{-\infty}^{\infty} e^{tx} f_X(x)dx$$

si X es una variable aleatoria absclutamente continua; y

$$m(t)=E(e^{tX})= \sum_{x} e^{tx} f_X(x)$$

si X es una variable aleatoria discreta.

Si la función generatriz de momentos existe, entonces es continuamente diferenciable para algún intervalo alrededor del origen. Si se diferencia r veces con respecto a t resulta

$$\frac{d^r}{dt^r}m(t)= \int_{-\infty}^{\infty} x^r e^{xt} f_X(x)dx,$$

si $t\to 0$ queda

$$\frac{d^r}{dt^r}m(0)=E[X^r]=\mu'_r,$$

en donde se observa que el r-ésimo momento se obtiene derivando r veces $m(t)$ evaluada cuando t=0. Por lo tanto los momentos de una distribución se obtienen derivando la función generatriz de momentos.

Ejemplos:

i) Sea X una variable aleatoria cuya función de densidad está definida por $f_X(x)=\lambda e^{-\lambda x}$ (Distribución Exponencial) entonces ¿Cuáles son su primer y segundo momento?.

$$m_X(t)=E[e^{tX}] = \int_0^\infty e^{tx}\lambda e^{-\lambda x}dx = \int_0^\infty \lambda e^{tx-\lambda x}\, dx = \int_0^\infty \lambda e^{-(\lambda-t)x}dx$$

$$= \frac{\lambda}{\lambda - t}\int_0^\infty (\lambda-t)e^{-(\lambda-t)x}dx = \frac{\lambda}{\lambda - t} \text{ para } t<\lambda$$

por lo tanto

$$m_x'(t)=\frac{\lambda}{(\lambda - t)^2} \text{ entonces } m'(0)=E[X]=1/\lambda$$

$$m_x''(t)= \frac{2\lambda}{(\lambda - t)^3} \text{ entonces } m''(0)=E[X^2]=2/\lambda^2$$

$$\text{var}[X]=E[X^2]-(E[X])^2=1/\lambda^2.$$

ii) Sea X una variable aleatoria cuya función de densidad esta definida por la función de densidad Poisson. ¿Cuáles son su primer y segundo momento?.

$$f_x(x,\lambda)=\begin{cases} \dfrac{e^{-\lambda}\lambda^x}{x!} & \text{para } x=0,1,2,\ldots \text{ y } \lambda>0 \\ 0 & \text{en otro caso} \end{cases}$$

$$m_x(t)=E[e^{tx}]=\sum_{x=0}^{\infty}\frac{e^{tx}\, e^{-\lambda}\, \lambda^x}{x!}$$

$$= e^{-\lambda}e^{\lambda e^t},$$

entonces

$$m_x'(t)=\lambda e^{-\lambda} e^{t} e^{\lambda e^{t}}$$

$$m_x''(t)=\lambda e^{-\lambda} e^{t} e^{\lambda e^{t}}[\lambda e^{t}+1]$$

de donde

$$E(X)=m_X'(0)=\lambda$$

y

$$\text{var}[X]=E[X^2]-(E[X])^2=\lambda(\lambda+1)-\lambda^2=\lambda.$$

Se plantea la pregunta de que si con la función generatriz de momentos, además de poder obtener los momentos, se podrá determinar la función de distribución de una variable aleatoria que se encuentra en término de otras variables que se conocen como se distribuyen. Para ello se analiza el siguiente ejemplo; considérese X variable aleatoria que se distribuye normalmente con media 0 y varianza 1. Sea $Y=X^2$, ¿Cuál es la función generatriz de momentos de Y?.

$$m_Y(t)=E[e^{tY}]=\int_{-\infty}^{\infty} e^{tx^2}\frac{1}{\sqrt{2\pi}}e^{-1/2x^2}dx$$

$$=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-1/2x^2(1-2t)}dx$$

$$=\frac{1}{(1-2t)^{-1/2}}\int_{-\infty}^{\infty}\frac{(1-2t)^{-1/2}}{\sqrt{2\pi}}e^{-1/2x^2(1-2t)}dx$$

la integral es igual a 1 porque es una normal $[0,1/(1-2t)]$

$$=(1-2t)^{1/2}=\left(\frac{1/2}{1/2-t}\right)^{1/2} \quad \text{para } t<1/2.$$

Por otro lado, se tiene que si $X \sim \Gamma(n,\lambda)$ entonces la función generatriz de momentos resulta:

$$\Gamma(n,\lambda) = \frac{\lambda^{n} e^{-\lambda x} x^{n-1}}{\Gamma(n)}$$

entonces

$$m_Y(t) = E[e^{tY}] = \int_0^{\infty} \frac{\lambda^{n} e^{tx} e^{-\lambda x} x^{n-1}}{\Gamma(n)} dx$$

$$= \frac{\lambda^{n}}{(\lambda - t)^{n}} \int_0^{-\infty} \frac{(\lambda - t)^{n}}{\Gamma(n)} e^{-(\lambda - t)x} x^{n-1} dx$$

$$= \left(\frac{\lambda}{\lambda - t} \right)^{n}$$

la cual se conoce como la función generatriz de momentos de la función de distribución gamma. En base a lo anterior, se podría decir que la variable aleatoria $Y=X^2$ se distribuye como una $\Gamma(1/2,1/2)$ por el hecho de tener igual función generatriz de momentos . Sin embargo, en general no es válido hacer este tipo de aseveraciones pues existen condiciones bajo las cuales una sucesión de momentos no determinan una única función de distribución. El problema de si una función de distribución es determinada ó no por su sucesión de momentos es referido como el Problema de Momentos [2] que no será discutido en este trabajo.

[2] Mood.- Introduction to the Theory of Statistics . Pág. 81.

II. Vectores Aleatorios.

Definición 5.6.

Sea $(\Omega\ A.P)$ un espacio de probabilidad y $X_1, X_2, \ldots, X_n$ variables aleatorias definidas sobre ese mismo espacio de probabilidad, $g(X_1, X_2, \ldots X_n)$ una función cualesquiera del vector aleatorio $(X_1, X_2, \ldots X_n)$ entonces

$$E(g(X_1, X_2, \ldots, X_n)) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(X_1, X_2, \ldots, X_n)$$

$$dF_{X_1, X_2, \ldots X_n}(X_1, X_2, \ldots, X_n)$$

siempre que $g(X_1, \ldots, X_n)$ sea integrable con respecto a $F_{X_1, \ldots, X_n}(x_1, \ldots, x_n)$.

Teorema 5.4.

Si $X_1, X_2, \ldots X_n$ son variables aleatorias mutuamente independientes entonces

$$E\, g_1(X_1) g_2(X_2) \ldots g_n(X_n) = \prod_{i=1}^{n} E\, g_i(X_i).$$

Demostración:

Si $X_1, X_2, \ldots X_n$ son mutuamente independientes entonces

$$F(x_1, x_2, \ldots x_n) = \prod_{i=1}^{n} F_i(x_i)$$

donde cada $F_i(x_i)$ es la función de distribución marginal de X_i, entonces

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g_1(x_1) g_2(x_2) \ldots g_n(x_n)\, dF(x_1, x_2, \ldots, x_n)$$

$$=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} g_1(x_1)g_2(x_2)\ldots g_n(x_n)\,dF_1(x_1)\,dF_2(x_2)\ldots dF_n(x_n)$$

$$=\prod_{i=1}^{n}\int_{-\infty}^{\infty} g_i(x_i)\,dF_i(x_i)=\prod_{i=1}^{n} E\, g_i(X_i)\ . \qquad \blacksquare$$

Esperanza Condicional.

El concepto de Esperanza Condicional se analiza primero para (X,Y) variables aleatorias posteriormente se presenta su generalización.

<u>Definición 5.7.</u>

Sea (X,Y) variables aleatorias y $g(.,.)$ una función cualesquiera de (x,y); entonces la *Esperanza Condicional* de $g(X,Y)$ dado que X=x se denota como $E[g(X,Y)|X=x]$, y se define como:

$$E[g(X,Y)|X=x] = \int_{-\infty}^{\infty} g(x,y)f_{Y|X}(y|x)\,dy$$

si (X,Y) son variables aleatorias continuas

$$E[g(X,Y)|X=x = \sum_j g(x,y_j)f_{Y|X}(y_j|x)$$

si (X,Y) son variables aleatorias discretas.

En el caso que se tenga $(X_1,\ldots,X_k,Y_1,\ldots,Y_m)$ un vector aleatorio de dimensión k+m entonces la esperanza condicional se define como:

$$E[g(X_1,\ldots,X_k,Y_1,\ldots,Y_m)|x_1,\ldots,x_k] =$$

$$=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x_1,\ldots,x_k,y_1,\ldots,y_m)$$

$$f_{Y_1,\ldots Y_m\, X_1,\ldots,X_m}(y_1,\ldots,y_m\ x_1,\ldots,x_k)\,dy_1,\ldots,dy_m.$$

Teorema 5.5.

Sea (X,Y) variables aleatorias, y g una función de Y entonces

$$E[g(Y)] = E\left[E[g(Y)|X]\right]$$

y en particular

$$E(Y) = E\left[E[Y|X]\right].$$

Demostración:

Se ha visto que $E[g(Y)X]$ es una función de x entonces se puede definir a $E[g(Y)X] = h(X)$ entonces resulta que

$$\begin{aligned} E\left[E[g(Y)|X]\right] &= E[h(X)] \\ &= \int_{-\infty}^{\infty} h(x)f_X(x)dx = \int_{-\infty}^{\infty} E[g(Y)|X]f_X(x)dx \\ &= \int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty} g(y)f_{Y|X}(y|x)dy\right]f_X(x)dx \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(y)f_{Y,X}(y,x)dydx \\ &= E[g(Y)]. \end{aligned}$$

∎

La esperanza condicional es muy utilizada dentro de la Estadística Bayesiana.

Definición 5.8. (Varianza Condicional).

La varianza de Y dado que X=x se define como

$$\text{var}[Y|X=x] = E[Y^2|X=x]-(E[Y|X=x])^2.$$

Teorema 5.6.

$$\text{var}[Y] = E[\text{var}(Y|X)] + \text{var}(E[Y|X]).$$

Demostración:

$$E(\text{var}[Y|X]) = E[E(Y^2|X)] - E[(E(Y|X)^2]$$
$$= E(Y^2) - (E[Y])^2 - E[(E(Y|X)^2] + (E[Y])^2$$
$$= \text{var}[Y] - E[(E(Y|X)^2] + (E[(E(Y|X)])^2$$
$$= \text{var}[Y] - \text{var}\,[E(Y|X)]. \quad \blacksquare$$

Tanto la esperanza como la varianza condicional son utilizadas en demostraciones presentadas en los siguientes capítulos.

MOMENTOS.

El momento no central de $F_{X_1,\ldots,X_n}(x_1,\ldots,x_n)$ se define asignando a $g(X_1,\ldots,X_n) = X_1^k X_2^k \ldots X_n^k$ y se denota por $\mu_{k1,k2,\ldots,kn}$ es decir

$$E(X_1^k X_2^k \ldots X_n^k) = \mu_{k1,k2,\ldots,kn}.$$

El momento central de $F_{X_1,\ldots,X_n}(x_1,\ldots,x_n)$ se define como,

$$\mu_{k1,k2,\ldots,kn} = E((X_1-\mu_1)^{k1}(X_2-\mu_2)^{k2}\ldots(X_n-\mu_n)^{kn}).$$

Un momento se dice que es de orden r si $\sum_{j=1}^{n} k_j = r$. En el caso de que el momento central sea de orden 2 se presentan dos casos:

i) $k_j=2$ para cualquier $1 \le j < n$, y $k_i=0$ para cualquier $i \ne j$ entonces

$$E(X_j-\mu_j)^2 = \sigma_j^2,$$

ii) $k_i=k_j=1$ para $1 \le i < j < n$ y cero para cualquier otra k's entonces

$$E(X_i-\mu_i)(X_j-\mu_j) = \mu_{0,0,\ldots 0,1,0,\ldots 1,0}$$

que es conocido como la *covarianza* entre X_i y X_j.

Definición 5.9.

Sea X y Y variables aleatorias entonces

$$\text{Cov }(X,Y)=E\ (X-E(X))(Y-E(Y))=E(XY)-(E(X))(E(Y))$$

si X y Y son independientes entonces $\text{Cov}(X,Y)=0$, sin embargo el inverso no es verdadero.

Ejemplo:

Sea (X,Y) un vector aleatorio discreto con función de distribucion $p(x,y)$ definida por:

$$p(x,y)=\begin{cases} p(-1,1)=p(-1,-1)=p(1,-1)=p(1,1)=1/8 \\ p(0,0)=1/2 \\ p(x,y)=0 \quad \text{en otro caso} \end{cases}$$

entonces

$$E(X)=E(Y)=0 \text{ y } \text{Cov}(X,Y)=E(XY)-E(X)E(Y)=0,$$

por otro lado se calculan las funciones de densidad marginal

$$p_x(1)=p_x(-1)=p_y(1)=p_y(-1)=1/4 \quad y \quad p_x(0)=p_y(0)=1/2$$

por lo tanto

$$p(0,0)\neq p_x(0)p_y(0)$$

lo cual implica que X y Y no son independientes.

Teorema 5.7. (Suma de Varianza).

Si X y Y son dos variables aleatorias con segundo momento finito entonces

$$\text{Var}(X+Y)=\text{Var}(X)+\text{Var}(Y)+2\text{Cov}(XY).$$

Demostración:

Si se desarrolla Var(X+Y) resulta que

$$\mathrm{Var}(X+Y)=E\,[(X+Y)-E(X+Y)]^2.$$

$$=\mathrm{Var}(X)+Var(Y)+2E[(X-E(X))(Y-E(Y))]. \blacksquare$$

Proposición 5.3.

Si $X_1, X_2, \ldots X_n$ son variables aleatorias con segundo orden finito entonces

$$\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right)=\sum_{i=1}^{n}\mathrm{Var}\,X_i+2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\mathrm{Cov}\,(X_i,Y_j).$$

Demostracion:

$$\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right)=E\left[\left(\sum_{i=1}^{n} X_i-E\left[\sum_{i=1}^{n} X_i\right]\right)^2\right]$$

$$=E\left[\sum_{i=1}^{n}\sum_{j=1}^{n}(X_i-E[X_i])(X_j-E[X_j])\right]$$

$$=\sum_{i=1}^{n}\sum_{j=1}^{n}E\left[(X_i-E[X_i])(X_j-E[X_j])\right]$$

$$=\sum_{i=1}^{n}\mathrm{Var}\,X_i+\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\mathrm{Cov}(X_i,X_j). \qquad \blacksquare$$

<u>Teorema 5.8.</u>

Si X y Y son dos variables aleatorias independientes con segundo momento finito entonces

$$\mathrm{Var}(X+Y)=\mathrm{Var}(X)+\mathrm{Var}(Y).$$

La demostración se ve claramente por el hecho del teorema 5.6 y por el hecho de que X y Y son independientes lo cual implica que $\mathrm{Cov}(X,Y)=0$. $\blacksquare$

El resultado del Teorema 5.8 se puede generalizar para $X_1, X_2, \ldots X_n$ variables aleatorias,

Proposición 5.4.

Si $X_1, X_2, \ldots X_n$ son variables aleatorias independientes entonces

$$\operatorname{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \operatorname{Var} X_i.$$

Teorema 5.9.

Sea $X_1, \ldots, X_n$ variables aleatorias si $Y=\Sigma_{i=1}^{n} \alpha_i X_i$ con $\alpha_1, \ldots, \alpha_n$ constantes entonces

$$E(Y) = \Sigma_{i=1}^{n} \alpha_i \mu_i$$

y

$$\operatorname{Var}(Y) = \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j \operatorname{Cov}[X_i, X_j] \geq 0.$$

Demostración:

$$E(Y) = E\left\{\sum_{i=1}^{n} \alpha_i X_i\right\} = \sum_{i=1}^{n} \alpha_i E(X_i) = \Sigma_{i=1}^{n} \alpha_i \mu_i .$$

y

$$\begin{aligned}
\operatorname{Var}(Y) = E((Y-E(Y))^2) &= E\left\{\left(\sum_{i=1}^{n} \alpha_i X_i - \sum_{i=1}^{n} \alpha_i \mu_i\right)^2\right\} \\
&= E\left\{\left(\sum_{i=1}^{n} \alpha_i (X_i - \mu_i)\right)^2\right\} \\
&= E\left\{\sum_{i=1}^{n} \alpha_i (X_i - \mu_i) \sum_{j=1}^{n} \alpha_j (X_j - \mu_j)\right\} \\
&= E\left\{\sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j (X_i - \mu_i)(X_j - \mu_j)\right\} \\
&= \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j \operatorname{Cov}[X_i, X_j],
\end{aligned}$$

sea $\operatorname{Cov}[X_i, X_j] = b_{ij}$ resulta que

$$\begin{vmatrix} b_{11} & \cdots & b_{1k} \\ b_{21} & \cdots & b_{2k} \\ \vdots & & \vdots \\ b_{k1} & \cdots & b_{kk} \end{vmatrix} \geq 0 \quad \text{para } k=1,\ldots,n$$

a esta matriz se le conoce como *matriz de covarianza*. Esta matriz tiene las siguientes propiedades,

i) la diagonal de $(b_{ij})_{nXn} = \text{var}(X_i)$,

ii) la matriz es simétrica y positiva,

iii) la $\text{Cov}(X_i, X_i) = \text{Var}(X_i)$.

<u>Teorema 5.10.</u>

Sea U, V dos funciones lineales de las variables aleatorias $X_1, \ldots, X_n$ es decir $U = \Sigma_{i=1}^{n} \alpha_i X_i$ y $V = \Sigma_{i=1}^{n} \beta_i X_i$ entonces

$$\text{Cov}\,(U,V) = \Sigma_{i=1}^{n} \Sigma_{j=1}^{n} \alpha_i \beta_i \text{Cov}(X_i, X_j).$$

Demostración:

$$E((U-E(U))(V-E(V))) =$$

$$= E\left\{ \left(\sum_{i=1}^{n} \alpha_i X_i - \sum_{i=1}^{n} \alpha_i \mu_i \right) \left(\sum_{j=1}^{n} \beta_j X_j - \sum_{j=1}^{n} \beta_j \mu_j \right) \right\}$$

$$= E\left\{ \sum_{i=1}^{n} \alpha_i (X_i - \mu_i) \sum_{j=1}^{n} \beta_j (X_j - \mu_j) \right\}$$

$$= E\left\{ \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \beta_j (X_i - \mu_i)(X_j - \mu_j) \right\}$$

$$= \Sigma_{i=1}^{n} \Sigma_{j=1}^{n} \alpha_i \beta_i \text{Cov}(X_i, X_j). \quad \blacksquare$$

Proposición 5.5.

Si $U = aX + b$ y $V = cY + d$ entonces $\text{Cov}(U,V) = ac\text{Cov}(X,Y)$.

Demostración:

$$\text{Cov}(U,V) = E((U-E(U))(V-E(V)))$$

$$= E\Big[[(ax+b) - E(ax+b)]\,[(cy+d) - E(cy+d)] \Big]$$

$$= E\Big[\,(ax+b)-a(E(x)+b)\;\;(cy+d)-c(E(y)+d)\,\Big]$$

$$= E\Big[\,a(x-E(x)\;\;c(y-E(y)\,\Big]$$

$$= ac\mathrm{Cov}(X,Y). \qquad \blacksquare$$

Coeficiente de Correlación.

Definición 5.10.

Sean X y Y dos variables aleatorias entonces el coeficiente de correlación entre X y Y se define como:

$$\rho_{XY}=\frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}X\ \mathrm{Var}Y}}.$$

Teorema 5.11.

Si U=ax+b y V=cy+d entonces $\rho_{UV}=\rho_{XY}$.

Demostración:

$$\rho_{UV}=\frac{\mathrm{Cov}(U,V)}{\sqrt{\mathrm{Var}U\ \mathrm{Var}V}}$$ aplicando la proposición 5.5 resulta

$$=\frac{ac\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(ax+b)}\ \sqrt{\mathrm{Var}(cy+d)}}$$

$$=\frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}X\ \mathrm{Var}Y}}=\rho_{XY}. \qquad \blacksquare$$

Proposición 5.6.

El coeficiente de correlación cumple con las siguientes propiedades:

i) si X,Y son independientes entonces $\rho_{XY} = 0$,

ii) si $\rho_{XY}=0$ no implica que X,Y sean independientes,

iii) $-1\leq \rho_{XY} \leq 1$,

iv) si $\rho_{XY}=-1$ ó 1 entonces implica que X,Y estan relacionadas linealmente.

Demostración:

i) Por definición se conoce que si X,Y son independientes entonces $\mathrm{Cov}(X,Y)=0$ por lo tanto $\rho_{XY}=0$

ii) Considere X,Y variables aleatorias tales que $Y=X^2$, con X simétrica con respecto al origen, y satisface la condición de que $E(X^4)$ es finita. Entonces la $E(X)=0$, y $E(XY)=E(X^3)=0$ claramente se observa que $\mathrm{Cov}(X,Y)=0$ por lo que $\rho_{XY}=0$, sin embargo X y Y son dependientes.

iii) Sea

$$U=\frac{X-\mu_x}{\sigma_x} \quad , \quad V=\frac{Y-\mu_y}{\sigma_y}$$

entonces $E(U)=E(V)=0$ y $\mathrm{Var}(U)=\mathrm{Var}(V)=1$. Por el teorema 5.11

$$\rho_{UV}=\rho_{XY}=\sigma_{UV}.$$

lo cual lleva a considerar que

$$0\leq\mathrm{Var}(X+Y)=\mathrm{Var}(X)+\mathrm{Var}(Y)+2\mathrm{Cov}(X,Y)=2+2\rho uv$$
$$0\leq\mathrm{Var}(X-Y)=\mathrm{Var}(X)+\mathrm{Var}(Y)-2\mathrm{Cov}(X,Y)=2-2\rho uv$$

de donde se ve que hay dos opciones para que se cumpla la desigualdad,

$$2+2\rho_{UV}\geq 0 \quad ó \ \rho_{UV}\geq -1$$
$$2-2\rho_{UV}\geq 0 \quad ó \ \rho_{UV}\leq 1$$

por lo tanto $-1\leq\rho_{UV}\leq 1$.

iv) Para demostrar esta propiedad vamos hacer uso de que

$$\mathrm{Var}\left(\frac{X}{\sqrt{\mathrm{Var}X}} \pm \frac{Y}{\sqrt{\mathrm{Var}Y}}\right) = 2 \pm 2\rho_{XY} \geq 0$$

primero se demuestra esta igualdad

$$\mathrm{Var}\left(\frac{X}{\sqrt{\mathrm{Var}X}} + \frac{Y}{\sqrt{\mathrm{Var}Y}}\right) =$$

$$= \mathrm{Var}\left(\frac{X}{\sqrt{\mathrm{Var}X}}\right) + \mathrm{Var}\left(\frac{Y}{\sqrt{\mathrm{Var}Y}}\right)$$

$$+ 2\,\mathrm{Cov}\left(\frac{XY}{\sqrt{\mathrm{Var}X\ \mathrm{Var}Y}}\right)$$

$$= E\left(\frac{X}{\sqrt{\mathrm{Var}X}}\right)^2 - \left[E\left(\frac{X}{\sqrt{\mathrm{Var}X}}\right)\right]^2 + E\left(\frac{Y}{\sqrt{\mathrm{Var}Y}}\right)^2$$

$$- \left[E\left(\frac{Y}{\sqrt{\mathrm{Var}Y}}\right)\right]^2 + 2\left[E\left(\frac{XY}{\sqrt{\mathrm{Var}X}\sqrt{\mathrm{Var}Y}}\right) - E\left(\frac{X}{\sqrt{\mathrm{Var}X}}\right)\right.$$

$$\left. E\left(\frac{Y}{\sqrt{\mathrm{Var}Y}}\right)\right] = 1 + 1 + 2\rho_{XY},$$

en el caso de $\mathrm{Var}\left(\frac{X}{\sqrt{\mathrm{Var}X}} - \frac{Y}{\sqrt{\mathrm{Var}Y}}\right)$ se realiza el mismo mismo procedimiento; por lo tanto $-1 \leq \rho_{XY} \leq 1$. Para que el coeficiente de correlación sea igual a uno, tendria que suceder que $\mathrm{Var}\left(\frac{X}{\sqrt{\mathrm{Var}X}} - \frac{Y}{\sqrt{\mathrm{Var}Y}}\right) = 0$ y esto ocurre solo si la variable toma el valor de una constante con probabilidad de 1, es decir que $\dfrac{X}{\sqrt{\mathrm{Var}X}} - \dfrac{Y}{\sqrt{\mathrm{Var}Y}} = c$, por lo tanto $X = \dfrac{\sqrt{\mathrm{Var}X}}{\sqrt{\mathrm{Var}Y}}\, Y + d$, donde $d = c\sqrt{\mathrm{Var}X}$. Se utiliza el mismo razonamiento en el caso de $\rho_{XY}=-1$, en donde

$$X = -\frac{\sqrt{\mathrm{Var}X}}{\sqrt{\mathrm{Var}Y}}\, Y + d. \qquad \blacksquare$$

Como se puede observar el coeficiente de correlación es una medida de la relación lineal entre las variable aleatorias y tiene un uso ilimitado dentro de la Estadística. Sin embargo hay que tener cuidado al usarlo debido a que establece la relación lineal entre las variables más no una relación de causalidad; por ejemplo, sea X que represente el número de iglesias en un pueblo y Y el número de cantinas en ese mismo pueblo, puede ocurrir que $\rho_{XY}=1$ sin embargo no hay razón para pensar que una dependa de la otra.

Función Generatriz de Momentos.

Los momentos de las funciones de densidad tienen un papel importante dentro de la Estadística. Por lo que, es conveniente tener una función que genere a todos los momentos.

<u>Definición 5.11.</u>

La *Función Conjunta Generatriz de Momentos* de $X_1, \ldots X_n$ variables aleatorias se define como:

$$m_{X_1,\ldots X_k}(t_1,\ldots t_k) = E\left[\exp \sum_{j=1}^{k} t_j X_j\right]$$

siempre y cuando la esperanza exista para todos los valores de $t_1,\ldots,t_k$ tales que $-h<t_j\leq h$ para algún $h>0$, $j=1,\ldots,k$.

El r-ésimo momento se puede obtener derivando r veces con respecto a t y sacando el límite de todas las t's cercanas a cero.

Propiedades:

i) Si $(X_i)_{i=1}^{n}$ son variables aleatorias mutuamente independientes entonces

$$m_{x1,\ldots,xn}(t_1,\ldots,t_n) = \prod_{j=1}^{n} m_i(t_i)$$

Demostración:

$$m_{x1,\ldots,xn}(t_1,\ldots,t_n) = E\left[\exp\sum_{j=1}^{n} t_j X_j\right] = E\left[\prod_{j=1}^{n} \exp t_j X_j\right] =$$

$$= \prod_{j=1}^{n} E\left[\exp t_j X_j\right] = \prod_{i=1}^{n} m_i(t_i).$$

ii) $m_{xi}(t_i) = m_{x1,\ldots,xn}(0,..,t_i,..,0)$

iii) $\dfrac{\delta^n}{\delta t_1\ldots\delta t_n} m_{x1,\ldots,xn}(t_1,\ldots,t_n)_{t_1,t_2,..,t_n=0} = E(X_1,\ldots,X_n)$

Ejemplo:

Sea X y Y variables aleatorias que se distribuyen normalmente ¿Cuál es la función generatriz de momentos de (X,Y)?.

$$m_{xy}(t_1,t_2) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{t_1X+t_2Y}\frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}}$$

$$\exp\left\{\frac{-1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2 - 2\rho\frac{x-\mu_x}{\sigma_x}\frac{y-\mu_y}{\sigma y} + \left(\frac{y-\mu_y}{\sigma_y}\right)^2\right]\right\}dxdy$$

si se hace

$$u = \frac{x-\mu_x}{\sigma_x} \qquad y \qquad v = \frac{y-\mu_y}{\sigma_y}$$

$$m_{xy}(t_1,t_2) = e^{t_1\mu_x+t_2\mu_y} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{t_1\sigma_x u+t_2\sigma_y v} \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}}$$

$$\exp\left\{-\frac{1}{2(1-\rho^2)}\left[u^2-2\rho uv+v^2\right]\right\} du dv$$

la exponencial dentro de la integral puede reexpresarse como

$$\exp\left\{-\frac{1}{2(1-\rho^2)}\left[(u^2-2\rho uv+v^2)-2(1-\rho^2)t_1\sigma_x u-2(1\ \rho^2)t_2\sigma_y v\right]\right\}$$

y si se completa el cuadrado para u y v la exponencial puede ser expresada como

$$\exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(u-2\rho uv-(1-\rho^2)t_1\sigma_x\right)^2-(1-\rho^2)(v-\rho t_1\sigma_x-t_2\sigma_y)^2\right.\right.$$

$$\left.\left.-(1-\rho^2)(t_1^2\sigma_x^2+2\rho t_1t_2\sigma_x\sigma_y+t_2^2\sigma_y^2)\right]\right\}$$

si se sustituye $w = \dfrac{u-\rho v-(1-\rho^2)t_1\sigma_x}{\sqrt{1-\rho^2}}$ y $z = v-\rho t_1\sigma_x-t_2\sigma_y$

se tiene que la exponencial puede escribirse como

$$\exp\left\{-\frac{1}{2}w^2-\frac{1}{2}z^2+(t_1^2\sigma_x^2+2\rho t_1t_2\sigma_x\sigma_y+t_2^2\sigma_y^2)\right\}$$

por lo tanto

$$m_{xy}(t_1,t_2)=e^{t_1\mu_x+t_2\mu_y}\exp\left\{\frac{1}{2}(t_1^2\sigma_x^2+2\rho t_1t_2\sigma_x\sigma_y+t_2^2\sigma_y^2)\right\}$$

$$\int_{-\infty}^{\infty}\mu\int_{-\infty}^{\infty}\frac{1}{2\pi}e^{-w^2/2-z^2/2}\,dw\,dz$$

entonces

$$m_{xy}(t_1,t_2) = \exp\left\{ t_1\mu_x + t_2\mu_y + \frac{1}{2}(t_1^2\sigma_x^2 + 2\rho t_1 t_2 \sigma_x\sigma_y + t_2^2\sigma_y^2)\right\}$$

ya que la doble integral es igual a uno.

$$\frac{\delta m}{\delta t_1}\ t_1,t_2{=}0 = \mu_x$$

$$\frac{\delta^2 m}{\delta t_1}\ t_1,t_2{=}0 = \mu_x^2 + \sigma_x^2$$

en el caso de que la diferenciación se realize sobre t_2 se obtiene

$$\frac{\delta m}{\delta t_2}\ t_1,t_2{=}0 = \mu_y$$

$$\frac{\delta^2 m}{\delta t_2}\ t_1,t_2{=}0 = \mu_y^2 + \sigma_y^2.$$

La función de distribución conjunta de $X_1,...,X_n$, no se puede determinar por medio de La función generatriz de momentos n-dimensional. (Ver Problema de Momentos).

FUNCION CARACTERISTICA

En este capitulo se analiza otro de los métodos utilizados para determinar la función de distribución de variables aleatorias definidas en términos de otras variables en las que se conoce como se distribuyen. Para ello, se estudia la Función Caracteristica, la Fórmula de Inversión, y el Teorema de Unicidad. Primero se analizan para variables aleatorias, y posteriormente se presenta su generalización a vectores aleatorios n-dimensionales.

Sea (Ω, A, P) una espacio de probabilidad, X una variable aleatoria definida sobre ese mismo espacio entonces;

Definición 6.1.

La *función caracteristica* de una variable aleatoria X se define como

$$\varphi_X(t)=E[e^{itX}] \qquad -\infty<t<\infty$$

donde $i=\sqrt{-1}$; por lo tanto si X es una variable aleatoria discreta

$$\varphi_X(t)=\sum e^{itx}f(x),$$

y si X es una variable aleatoria continua y la integral existe entonces,

$$\varphi_X(t)=\int_{-\infty}^{\infty} e^{itx}dF(x).$$

Antes de continuar con el análisis de la función caracteristica se enuncian algunas propiedades de los números complejos:

i) Cualquier número complejo z puede ser escrito de la forma $z=x+iy$ donde x y y son números reales, *ii*) El módulo de $z=x+iy$ esta definido por $z=(x^2+y^2)^{1/2}$,

iii) La distancia entre dos números complejos z_1 y z_2 esta definida por $|z_1-z_2|$,

iv) $E(Z)=E(X+iY)=E(X)+iE(Y)$ donde $E(X)$ y $E(Y)$ estan bien definidas.

Además, considérese que si $z=it$, donde t es un número real, entonces

$$e^z = \sum_{n=0}^{\infty}\frac{(it)^n}{n!}$$

$$= \left(1+it-\frac{t^2}{2}-\frac{it^3}{3!}+\frac{t^4}{4!}+\frac{it^5}{5!}-\ldots\right)$$

$$= \left(1-\frac{t^2}{2!}+\frac{t^4}{4!}-\ldots\right)+i\left(t-\frac{t^3}{3!}+\frac{t^5}{5!}-\ldots\right)$$

por lo tanto

$$e^{it}=\cos t + i\operatorname{sen} t \tag{1}$$

haciendo uso de que $\cos(-t)=\cos(t)$ y de que $\operatorname{sen}(-t)=-\operatorname{sen}t$ resulta que

$$e^{-it}=\cos t - i\operatorname{sen} t \tag{2}$$

entonces resolviendo el sistema de ecuaciones formado por (1) y (2) se tiene que,

$$\cos t = \frac{e^{it}+e^{-it}}{2} \quad y \quad \operatorname{sen} t = \frac{e^{it}-e^{-it}}{2i};$$

además

$$e^{it} = (\cos^2 t + \operatorname{sen}^2 t)^{1/2} = 1. \quad \blacksquare$$

Continuando con el estudio de la función característica, se observa que en caso de que X sea una variable aleatoria y t una constante entonces $e^{itX}=1$ por lo que la función característica es acotada, esto es:

$$\varphi_X(t) = E[e^{itX}] \le E\left[e^{itX}\right]=E(1)=1.$$

El siguiente teorema establece la función característica de una variable aleatoria expresada como una relación lineal, es decir,

<u>Teorema 6.1.</u>

Sea X, Y variables aleatorias tales que $Y=aX+b$ donde a y b son constantes entonces

$$\varphi_Y(t)= \varphi_X(at)\,e^{ibt}.$$

Demostración:

$$\varphi_Y(t)=E[e^{itY}]=E[e^{it(ax+b)}]=e^{itb}E(e^{itax})=e^{itb}\varphi(at). \qquad \blacksquare$$

Ejemplo:

Sea X variable aleatoria con función de distribución normal $X\text{-}N(\mu,\sigma^2)$. ¿Cuál es la función característica Y=2X?.

La función característica de la variable X se define como:

$$\varphi_Y(t)= \varphi_X(2t)$$

entonces

$$\varphi_X(t) = \int_{-\infty}^{\infty} \frac{e^{itx}}{\sqrt{2\pi}\,\sigma} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] dx$$

sea $y = [(x-\mu)/\sigma]$ entonces

$$\varphi_X(t) = \int_{-\infty}^{\infty} \frac{\exp\,[it(\sigma y + \mu)]}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2}\right)dy$$

$$= e^{it\mu} \int_{-\infty}^{\infty} \frac{\exp[(1/2)(y^2 - 2it\sigma y)]}{\sqrt{2\pi}}\, dy$$

$$= e^{it\mu} \int_{-\infty}^{\infty} \frac{\exp[(1/2)(y - it\sigma)^2 - (1/2)(i^2 t^2 \sigma^2)]}{\sqrt{2\pi}}\, dy$$

$$= \exp\left(it\mu - \frac{t^2\sigma^2}{2}\right) \int_{-\infty}^{\infty} \frac{\exp[-(1/2)(y - it\sigma)^2]}{\sqrt{2\pi}}\, dy$$

la integral es igual a uno, ya que se trata de la distribución normal, entonces resulta que

$$\varphi_X(t) = \exp\left(it\mu - \frac{t^2\sigma^2}{2}\right)$$

por lo tanto

$$\varphi_Y(t) = \exp\left(i2t\mu - 2t^2\delta^2\right)$$

Teorema 6.2.

Sea X y Y variables aleatorias independientes tales que $Z=X+Y$ entonces la función característica de Z se define como:

$$\varphi(Z)=\varphi(Y)\varphi(X).$$

Demostración:

$$\varphi(Z)=E[e^{itz}]=E[e^{it(x+y)}]=E(e^{itx})E(e^{ity})=\varphi(X)\varphi(Y). \quad \blacksquare$$

Ejemplo:

Sean X y Y variables aleatorias independientes distribuidas normalmente, entonces ¿Cuál es la función característica de $Z=X+Y$?.

Si $E(X)=\mu_1$, $\mathrm{Var}(X)=\sigma_1^2$, $E(Y)=\mu_2$, $\mathrm{Var}(Y)=\sigma_2^2$ entonces la funcion característica de las variables X y Y esta dada por:

$$\varphi_X(t)=\exp\left(it\mu_1-\frac{t^2\sigma_1^2}{2}\right) \qquad \varphi_Y(t)=\exp\left(it\mu_2-\frac{t^2\sigma_2^2}{2}\right)$$

respectivamente, por lo tanto la función característica de la suma es

$$\varphi_Z(t)=\varphi_X(t)\varphi_Y(t)=\exp\left(it(\mu_1+\mu_2)-\frac{1}{2}\left(\sigma_1^2+\sigma_2^2\right)t^2\right).$$

Por otro lado se tiene que, los momentos de una variable aleatoria, se pueden expresar por medio de la función característica de la misma. Es decir, si se deriva la función característica $\varphi_X(t)$ con respecto a t se obtiene,

$$\varphi_X'(t)=i\int_{-\infty}^{\infty}xe^{itx}dF(x),$$

$$\varphi_X''(t)=i^2\int_{-\infty}^{\infty}x^2e^{itx}dF(x),$$

y en general

$$\varphi_X^k(t)=i^k\int_{-\infty}^{\infty}x^ke^{itx}dF(x) \qquad (t=1,2,...)$$

para t=0 las integrales en los segundos miembros de las fórmulas obtenidas representan los momentos de la variable aleatoria X. Por lo tanto si se supone que t=0, se obtiene que

$$\varphi^k(0)=i^kE(x^k)$$

por lo tanto

$$E(x)=\frac{1}{i}\varphi'(0)$$

$$E(x^2)=-\varphi''(0)$$

estos resultados se pueden obtener siempre y cuando el momento correspondiente de la variable aleatoria X exista.

Ejemplo:

1) Sea X una variable aleatoria con función de distribución Poisson entonces, ¿Cuál es su función caracteristica?.

$$\varphi(t)=\sum_{x=0}^{\infty} e^{itx}\frac{e^{-\lambda}\lambda^{x}}{x!}=\sum_{x=0}^{\infty}\frac{(\lambda e^{it})^{x}e^{-\lambda}}{x!}$$

donde

$$\varphi(t)=e^{-\lambda}e^{\lambda e^{it}}$$

$$\varphi'(t)=e^{-\lambda}e^{\lambda e^{it}}\, i\lambda e^{it}$$

$$\varphi''(t)=e^{-\lambda}e^{\lambda e^{it}}\, i^2\lambda^2 e^{2it}+e^{-\lambda}e^{\lambda e^{it}}\, i^2\lambda e^{it}$$

por lo tanto si se evalúa en cero $\varphi'(0)=i\lambda$, y $\varphi''(0)=i^2\lambda^2+i^2\lambda$ entonces $E(x)=\lambda$ y $Var(x)=\lambda$.

2) Sea X una variable aleatoria con función de distribución Gamma $X\sim\Gamma(r,\lambda)$ entonces ¿Cuál es su función característica?.

$$\varphi_X(t)=\int_0^{\infty}\frac{\lambda^r}{\Gamma(r)}e^{itx}x^{r-1}e^{-\lambda r}dx$$

$$=\frac{\lambda^r}{(\lambda-it)^r}\int_0^{\infty}\frac{(\lambda-it)^r}{\Gamma(r)}x^{r-1}e^{-(\lambda-it)^r}dx$$

$$=\frac{\lambda^r}{(\lambda-it)^r}$$

se evalua la primera y segunda derivada

$$\varphi_X'(t)=r\lambda^r(\lambda-it)^{-r-1}i$$

$$\varphi_X''(t)=r(r+1)\lambda^r(\lambda-it)^{-r-2}i^2$$

entonces si t=0 se tiene que

$$\varphi_X'(0) = \frac{ir}{\lambda} \quad \varphi_X''(0) = \frac{r(r+1)i^2}{\lambda^2}$$

por lo tanto

$$E(X) = \frac{r}{\lambda} \; ; \; y \quad Var(X) = \frac{r(r+1)}{\lambda^2} - \frac{r^2}{\lambda^2} = \frac{r}{\lambda^2} \, .$$

Se ha visto que la primera y segunda derivada de la función característica evaluada en cero proporcionan los dos primeros momentos de las funciones de distribución; sin embargo está no es la única cualidad de la función caracteristica, puesto que también puede ser utilizada para determinar la función de distribución de una variable aleatoria expresada en término de otra que se conoce su distribución. Para ello se enuncian el Teorema de Unicidad y la Fórmula de Inversión.

Teorema 6.3. (Teorema de Unicidad).

Una función de distribución esta determinada unicamente por su función caracteristica y viceversa.

El Teorema de Unicidad establece que dada una función $\varphi(t)$, si esta es la función característica, se puede determinar $F(x)$ apartir de ella. Con esto se puede concluir que las propiedades de la función de distribución pueden ser estudiadas a través de las correspondientes propiedades de la función caracteristica. El siguiente teorema establece el proceso mediante el cual se determina la función de distribución apartir de la función característica.

Teorema 6.4.

Sean $\varphi(t)$ y $F(t)$ las funciones caracteristica y de distribución respectivamente de una variable aleatoria X. Si x_1 y x_2 son puntos de continuidad de la función $F(x)$, entonces,

$$F(x_2)-F(x_1)=\frac{1}{2\pi}\lim_{c\to\infty}\int_{-c}^{c}\frac{e^{-itx_1}-e^{-itx_2}}{it}\varphi(t)dt.$$

Demostración:

Sustituyendo $\varphi(t)$ resulta

$$J=\frac{1}{2\pi}\int_{-c}^{c}\int_{-\infty}^{\infty}\frac{1}{it}\left[e^{it(z-x_1)}-e^{it(z-x_2)}\right]dF(z)dt$$

el orden de integración en la última integral puede ser cambiado, ya que la integral con respecto a z converge absolutamente y la integral con respecto a t tiene límites de integración finitos.

$$J=\frac{1}{2\pi}\int_{-\infty}^{\infty}\left[\int_{-c}^{c}\frac{e^{it(z-x_1)}-e^{it(z-x_2)}}{it}dt\right]dF(z)$$

$$=\frac{1}{2\pi}\int_{-\infty}^{\infty}\left[\int_{-c}^{c}\frac{e^{it(z-x_1)}-e^{it(z-x_2)}+e^{-it(z-x_1)}-e^{-it(z-x_2)}}{it}dt\right]dF(z)$$

$$=\frac{1}{\pi}\int_{-\infty}^{\infty}\int_{0}^{c}\left[\frac{\text{sen }t(z-x_1)}{t}-\frac{\text{sen }t(z-x_2)}{t}\right]dtdF(z)$$

de Cálculo se conoce que si $c\to\infty$,

$$\frac{1}{\pi}\int_{0}^{c}\frac{\text{sen }\alpha t}{t}dt\to\begin{cases}\frac{1}{2} & \text{si } \alpha>0\\ -\frac{1}{2} & \text{si } \alpha<0\end{cases}$$

converge uniformemente con respecto a x en cada intervalo $\alpha>\delta>0$ ó $\alpha<-\delta$ y que para $\alpha\leq\delta$ resulta que

$$\left| \frac{1}{\pi} \int_{o}^{c} \frac{\text{sen } \alpha t}{t} dt \right| < 1;$$

supóngase que $x_1<x_2$ y si se representa J en forma de una suma resulta

$$J= \int_{-\infty}^{x_1-\delta} + \int_{x_1-\delta}^{x_1+\delta} + \int_{x_1+\delta}^{x_2-\delta} + \int_{x_2-\delta}^{x_2+} + \int_{x_2+\delta}^{\infty} \varphi(x_1,x_2)dF(z)$$

en donde por notación resulta

$$\varphi(x_1,x_2)= \frac{1}{\pi} \int_{o}^{c} \left\{ \frac{\text{sen } t(z-x_1)}{t} - \frac{\text{sen } t(z-x_2)}{t} \right\} dt$$

se determina $\delta>0$, de tal forma que $x_1+\delta<x_2-\delta$. El intervalo $-\infty<z<x_1-\delta$ contiene las desigualdades $z-x_1<\delta$ y $z-x_2<-\delta$ por lo tanto

$$\int_{-\infty}^{x_1-\delta} \varphi(x_1,x_2)dF(z)\rightarrow 0 \text{ cuando } c\rightarrow\infty.$$

Análogamente en el caso de $x_2+\delta<z<\infty$ y $c\rightarrow\infty$.

$$\int_{x_2+\delta}^{\infty} \varphi(x_1,x_2)dF(z)\rightarrow 0 \text{ cuando } c\rightarrow\infty.$$

Ya que las desigualdades $z-x_1>\delta$ y $z-x_2<-\delta$ estan contenidas en el intervalo $x_1+\delta<z<x_2-\delta$ y $c\rightarrow\infty$, entonces

$$\int_{x_1-\delta}^{x_2+\delta} \varphi(x_1,x_2)dF(z)\rightarrow \int_{x_1+\delta}^{x_2-\delta} dF(z)=F(x_2-\delta)-F(x_1+\delta)$$

y

$$\left| \int_{x_1-\delta}^{x_1+\delta} \varphi(x_1,x_2)dF(z) \right| <2 \int_{x_1-\delta}^{x_1+\delta} dF(z)=2[F(x_1+\delta)-F(x_1-\delta)],$$

$$\left| \int_{x_2-\delta}^{x_2+\delta} \varphi(x_1,x_2)dF(z) \right| <2 \int_{x_2-\delta}^{x_2+\delta} dF(z)=2[F(x_1+\delta)-F(x_1-\delta)]$$

entonces para cualquier $\delta>0$

$$\overline{\lim_{C\to\infty}} J=F(x_2-\delta)-F(x_1+\delta)+R(x_1,x_2)$$

$$\underline{\lim}_{C\to\infty} J=F(x_2-\delta)-F(x_1+\delta)+R(x_1,x_2)$$

y como J no depende de δ, resulta que:

$$\lim_{C\to\infty} J=F(x_1)-F(x_2). \qquad \blacksquare$$

Se conoce como la Fórmula de Inversión.

Ejemplo:

Si X,Y son dos variables aleatorias independientes con función de distribución normal entonces la función característica de $Z=X+Y$ se define como

$$\varphi_Z(t)=\varphi_X(t)\varphi_Y(t)=\exp\left(it(\mu_1+\mu_2) - \frac{1}{2} \left(\sigma_1^2 + \sigma_2^2\right)t^2\right).$$

analizando se observa que es la función característica de una variable aleatoria W distribuida normalmente con parámetros $\mu=\mu_1+\mu_2$, y $\sigma^2=\sigma_1^2+\sigma_2^2$. Por el teorema de unicidad se concluye que Z variable aleatoria se distribuye normalmente. Se puede observar que este método es menos tedioso que la transformación de variables aleatorias con el que también se determinó la función de distribución de la suma de dos variables aleatorias independientes normalmente distribuidas.

FUNCION CARACTERISTICA DE UN VECTOR ALEATORIO.

En está sección se presenta la generalización de la función característica para vectores aleatorios n-dimensional; no se realiza la demostración de los conceptos debido a que se considera que son objeto de Análisis Númerico. Se estudian detalladamente dentro de las Transformaciones de Fourier.

<u>Definición 6.2.</u>

La función característica de un vector aleatorio $(X_1, X_2, \ldots, X_n)$ se define como

$$\varphi(t_1, t_2, \ldots, t_n) = E(\exp(i \sum_{k=1}^{n} t_k x_k))$$

donde $t_1, t_2, \ldots t_n$ son valores reales.

La definición anterior implica que la función característica del vector aleatorio puede ser definida por:

$$\varphi(t_1, t_2, \ldots, t_n) = \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} \exp\ (i(t_1x_1 + \ldots + t_nx_n))\ f(x_1, \ldots, x_n)dx_1 \ldots dx_n.$$

La función característica es uniformemente continua sobre todo el espacio $(-\infty<t_j<\infty,\ 1\leq j<n)$ y tiene las siguientes propiedades:

i) $\varphi(0,0,\ldots 0) = 1,$

ii) $|\varphi(t_1, t_2, \ldots t_n)| \leq 1 \qquad (-\infty<t_k<\infty,\ k=1,2,\ldots n).$

Si los componentes del vector aleatorio $(X_1, X_2, \ldots X_n)$ son independientes, entonces por definición la función característica es igual al producto de funciones características de sus componentes, es decir

$$\varphi(t_1, t_2, \ldots t_n) = \varphi_1(t_1)\varphi_2(t_2), \ldots, \varphi_n(t_n).$$

Teorema 6.5.

Si la función característica del vector aleatorio $(X_1,\ldots,X_n)$ es $\varphi(t_1,\ldots,t_n)$ entonces la función característica del vector aleatorio $(\sigma_1X_1+\alpha_1,\ldots,\sigma_nX_n+\alpha_n)$ donde cada σ_i $(1\le i\le n)$ son constantes reales esta definida por

$$\exp\left(i\sum_{k=1}^{n}\alpha_k t_k\right)\varphi(\sigma_1 t_1,\ldots,\sigma_n t_n).$$

Por otro lado, se tiene que la función generatriz de momentos del vector aleatorio se puede determinar por medio de la función característica ya que si se deriva $\varphi(t_1,\ldots t_n)$ k_1-veces con respecto a t_1, k_2 veces con respecto a t_2 y así sucesivamente resulta

$$\frac{d^{k_1+\ldots\ldots+k_n}}{dt^{k_1}\ldots\ldots d\,t^{k_n}}\varphi(t_1,\ldots,t_n) = \left(i^{k_1+\ldots+k_n}\right)\int_{-\infty}^{\infty}\ldots\int_{-\infty}^{\infty}x_1^{k_1}\ldots x_2^{k_n}$$

$$\exp\left\{i(t_1x_1+\ldots\ldots+t_nx_n)\right\}f(x_1,\ldots,x_n)\,dx_1,\ldots x_n$$

de donde si $t_1=t_2=\ldots t_n=0$ se obtiene

$$\frac{d^{k_1+\ldots\ldots+k_n}}{dt^{k_1}\ldots\ldots d_{\cdot}^{k_n}}\varphi(t_1,\ldots,t_n) = \left(i^{\sum_{j=1}^{n}k_j}\right)\int_{-\infty}^{\infty}\ldots\int_{-\infty}^{\infty}x_1^{k_1}\ldots x_2^{k_n}$$

$$f(x_1,\ldots,x_n)\,dx_1,\ldots x_n$$

lo cual implica que

$$E(X_1^{k_1},X_2^{k_2},\ldots,X_n^{k_n}) = i^{\left(-\sum_{j=1}k_j\right)}\frac{d^{k_1+\ldots\ldots+k_n}}{dt_1^{k_1}\ldots dt_n^{k_n}}\varphi(t_1,\ldots,t_n)$$

es decir

$$E(X_1^{k_1},X_2^{k_2},\ldots,X_n^{k_n}) = i^{\left(-\sum_{j=1}k_j\right)}\frac{d^{k_1+\ldots\ldots+k_n}}{dt_1^{k_1}\ldots dt_n^{k_n}}\varphi(0,\ldots\ldots,0). \blacksquare$$

Ejemplo:

1) Sea $(X_1,\ldots,X_n)$ un vector aleatorio con función de distribución multinomial ¿CuÁl es su función caracteristica?.

La función de distribución multinomial se define como

$$f_{X_1,\ldots,X_n}(x_1,\ldots,x_n) = \frac{N!}{x_1!,\ldots,x_n!} p_1^{x_1},\ldots,p_n^{x_n}$$

en donde N representa el número de repeticiones, entonces

$$\varphi(t_1,\ldots,t_n) =$$

$$= \sum_{x_1,\ldots,x_n} \frac{N!}{x_1!\ldots x_n!} \exp\,[i(t_1x_1+\ldots+t_nx_n)]\,p_1^{x_1}\ldots p_n^{x_n}$$

$$= \sum_{x_1,\ldots,x_n} \frac{N!}{x_1!\ldots x_n!} (p_1e^{it_1})^{x_1}\ldots(p_ne^{it_n})^{x_n}$$

por lo tanto

$$\varphi(t_1,\ldots,t_n) = (p_1e^{it_1}+\ldots+p_ne^{it_n})^N$$

y

$$\frac{\delta}{\delta t_j}\varphi(t_1,\ldots,t_n) = Np_j ie^{it_j}(p_1e^{it_1}+\ldots+p_ne^{it_n})^{N-1},$$

$$\frac{\delta^2}{\delta t_j\,\delta t_k}\varphi(t_1,\ldots,t_n) =$$

$$= \begin{cases} N(N-1)i^2p_jp_ke^{it_j+it_k}(p_1e^{it_1}+\ldots+p_ne^{it_n})^{N-2} & j\neq k \\ N(N-1)i^2p_j^{\,2}e^{2it_j}(p_1e^{it_1}+\ldots+p_ne^{it_n})^{N-2} & \\ \qquad + Np_ji^2e^{it_j}(p_1e^{it_1}+\ldots+p_ne^{it_n})^{N-1} & j=k \end{cases}$$

si se evalúa en $(0,\ldots0)$ resulta que $\mu_j = Np_j$, $j=1,2,\ldots,n$

$$\sigma_{jk}^2 = \begin{cases} N(N-1)p_j p_k - N^2 p_j p_k = -Np_j p_k & j \neq k \\ N(N-1)p_j^2 + Np_j - N^2 p_j^2 = Np_j(1-p_j) & j = k \end{cases}.$$

2) Sea (X_1, X_2) un vector aleatorio con función de distribución normal ¿Cuál es su función caracaterística?.

$$\varphi(t_1,t_2) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{\exp[i(t_1 x + t_2 y)]}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\right.$$

$$\left.\left[\frac{(x-\mu_x)^2}{\sigma_x^2} - \frac{2\rho}{\sigma_x\sigma_y}(x-\mu_x)(y-\mu_y) + \frac{(y-\mu_y)^2}{\sigma_y^2}\right]\right\} dx\,dy$$

sustituyendo $u = (x - \mu_x)/\sigma_x$ y $v = (y - \mu_y)/\sigma_y$ resulta que

$\varphi(t_1,t_2) =$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{\exp[i(t_1\sigma_x u + t_2\sigma_y v + t_1\mu_x + t_2\mu_y)]}{2\pi\sqrt{1-\rho^2}}$$

$$\exp\left[-\frac{1}{2(1-\rho^2)}(u^2 - 2\rho uv + v^2)\right] du\,dv$$

$$= \exp[i(t_1\mu_x + t_2\mu_y)] \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{\exp[i(t_1\sigma_x u + t_2\sigma_y v)]}{2\pi\sqrt{1-\rho^2}}$$

$$\exp\left[-\frac{1}{2(1-\rho^2)}(u^2 - 2\rho uv + v^2)\right] du\,dv$$

se completa el cuadrado como se hizo en el caso de variables aleatorias

$\varphi(t_1,t_2) =$

$$= \exp[i(t_1\mu_x + t_2\mu_y)]\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{\exp(it_1\sigma_x u)\exp\left[-\dfrac{(u-\rho v)^2}{2(1-\rho^2)}\right]}{\sqrt{2\pi}\sqrt{1-\rho^2}}\,du$$

$$\frac{\exp(it_2\sigma_y v)\exp(-v^2/2)}{\sqrt{2\pi}}\,dv$$

la primera integral es la función característica para variables aleatorias normalmente distribuidas, si se reemplaza μ por ρv, t por $t_1\sigma_x$ y σ^2 por $1-\rho^2$; por lo tanto queda

$$\varphi(t_1,t_2) = \exp[i(t_1\mu_x + t_2\mu_y)]\int_{-\infty}^{\infty}\exp\left[it_1\sigma_x\rho v - \frac{t_1^2\sigma_x^2(1-\rho^2)}{2}\right]$$

$$\frac{\exp(it_2\sigma_y v - (v^2/2))}{\sqrt{2\pi}}\,dv$$

$$= \exp\left[i(t_1\mu_x + t_2\mu_y) - \frac{t_1^2\sigma_x^2(1-\rho^2)}{2}\right]$$

$$\int_{-\infty}^{\infty}\frac{\exp(i(t_1\sigma_x\rho + t_2\sigma_y)v)}{\sqrt{2\pi}}\exp\left(-\frac{v^2}{2}\right)dv$$

$$= \exp\left[i(t_1\mu_x+t_2\mu_y) - \frac{t_1^2\sigma_x^2(1-\rho^2)}{2} - \frac{(t_1\sigma_x\rho + t_2\sigma_y)^2}{2}\right]$$

esto se puede escribir como

$$\varphi(t_1,t_2) = \exp\left[i(t_1\mu_x + t_2\mu_y) - (1/2)(t_1^2\sigma_x^2 + 2\rho\sigma_x\sigma_y t_1 t_2 + t_2^2\sigma_y^2\right]$$

evaluando la primera y segunda derivada se tiene

$$\frac{\delta}{\delta t_1}\varphi(t_1,t_2) = \varphi(t_1,t_2)[i\mu_x - \sigma_x^2 t_1 - \rho\sigma_x\sigma_y t_2]$$

$$\frac{\delta}{\delta t_2}\varphi(t_1,t_2) = \varphi(t_1,t_2)[i\mu_y - \sigma_y^2 t_2 - \rho\sigma_x\sigma_y t_1]$$

$$\frac{\delta^2}{\delta t_1^2}\varphi(t_1,t_2) = \varphi(t_1,t_2)(-\sigma_x^2) + [i\mu_x - \sigma_x^2 t_1 - \rho\sigma_x\sigma_y t_1]^2\varphi(t_1,t_2)$$

$$\frac{\delta^2}{\delta t_2^2}\varphi(t_1,t_2) = \varphi(t_1,t_2)(-\sigma_y^2) + [i\mu_y - \sigma_y^2 t_2 - \rho\sigma_x\sigma_y t_2]^2\varphi(t_1,t_2)$$

$$\frac{\delta^2}{\delta t_1^2\delta t_2^2}\varphi(t_1,t_2) = \varphi(t_1,t_2)(-\rho\sigma_x\sigma_y) + \varphi(t_1,t_2) + [i\mu_x - \sigma_x^2 t_1 - \rho\sigma_x\sigma_y t_1][i\mu_y - \sigma_y^2 t_2 - \rho\sigma_x\sigma_y t_2]$$

como $\varphi(0,0)=1$, entonces si se evalua cada una de las derivadas en $(0,0)$ se obtiene μ_x, μ_y, $\sigma_x^2+\mu_x^2$, $\sigma_y^2+\mu_y^2$, $\rho\sigma_x\sigma_y+\mu_x\mu_y$ respectivamente.

<u>Teorema 6.6.</u>

Si $\varphi(t_1,t_2,\ldots,t_n)$ es la función característica y $F(x_1,x_2,\ldots,x_n)$ es la función de distribución del vector aleatorio $(X_1,X_2,\ldots X_n)$ entonces

$$P(a_k \le x_k < b_k\ ,\ k=1,2,\ldots,n) = \lim_{T\to\infty}\frac{1}{(2\Pi)^n}\int_{-T}^{T}\int_{-T}^{T}\cdots\int_{-T}^{T}\prod_{i=1}^{n}\frac{e^{-it_k a_k} - e^{-it_k b_k}}{it_k}\varphi(t_1,\ldots,t_n)dt_1..dt_n\ .$$

donde a_k, b_k son números reales que satisfacen que la probabilidad del punto $(X_1,\ldots X_n)$ caiga dentro de la superficie del paralilepípedo $a_k \le x_k < b_k (k=1,2,\ldots n)$ es cero. A este Teorema se le conoce como la Fórmula de Inversión para vectores aleatorios.

Teorema 6.7. (Teorema de Unicidad).

La función de distribución $F(x_1, x_2, \ldots x_n)$ esta determinada unicamente por la función característica y viceversa.

Como se ha podido observar la función característica resulta una herramienta muy útil dentro de la probabilidad ya que permite determinar la función de distribución de cualquier variable aleatoria expresada en término de otras que se conoce como se distribuyen, así como los momentos de cualesquiera variables aleatorias en forma única. En particular, se puede utilizar para resolver problemas relacionados con la distribución de variables aleatorias independientes.

CONVERGENCIA

En este capítulo se analizan las Leyes de los Grandes Números y el Teorema del Límite Central. Con el primero se expone y aclara la "interpretación frecuentista de la probabilidad", y con el segundo se estudian las leyes que caracterizan a la suma de un número grande de variables aleatorias.

LEY DE LOS GRANDES NUMEROS.

Cuando se enunció la definición Frecuentista de Probabilidad en el Capítulo I, se estableció que la razón n_A/n, donde n_A es el número de veces que el evento A ocurre en n repeticiones, se aproxima a la probabilidad p, cuando n crece. Pero ¿Cómo garantizar que efectivamente el límite existe y es p?. A lo largo de la historia, ha existido la inquietud de diversos matemáticos para establecer las condiciones bajo las cuales las variables aleatorias cumplan con está interrogante.

A finales del siglo XVII y principios del siglo XVIII James Bernouilli demostró el teorema que actualmente lleva su nombre; dicho teorema fué publicado por primera vez en el tratado *Arc Conjectandi*. A principios del siglo XIX Poisson demostró un teorema similar pero con condiciones mucho más generales. Hasta mediados de ese mismo siglo ningún progreso se realizó, pero en el año de 1866 un matemático ruso, Tchebychev demostró el teorema que ahora lleva su nombre; años más tarde Markov observó que el resultado de Tchebychev presentaba la oportunidad de ser generalizado. En el año de 1926, Kolmogorov estableció las condiciones necesarias y suficientes para que una sucesión de variables aleatorias mutuamente independientes cumplieran con la Ley de los Grandes Números. En 1828, Kintchine demostró que si las variables aleatorias X_n no solo son independientes, sino también, son idénticamente distribuidas, entonces la existencia de la esperanza de X_n era una condición suficente para el cumplimiento de la Ley Fuerte de los Grandes Números.

Para introducirnos en el estudio de la Leyes de los Grandes Números, considerése que si (Ω, A, P) es un espacio de probabilidad, y $(X_n)_{n=1}^{\infty}$ una sucesión de variables aleatorias definidas sobre ese mismo espacio entonces:

i) La sucesión de variables aleatoria $(X_n)_{n=1}^{\infty}$ Converge en Probabilidad a X si para cada c>0 y para δ>0 existe una N tal que

$$P(|X_n-X|>c) \rightarrow \delta, \qquad \text{para toda } n>N \qquad (1)$$

lo anterior se denota como

$$\lim_{n\to\infty} P(|X_n-X|\leq c)=1 \text{ ó bien } X_n \xrightarrow{P} X.$$

ii) La sucesión de variables aleatorias $(X_n)_{n=1}^{\infty}$ Converge Casi Seguramente a X si para toda r>1 existe N tal que

$$P\left\{X_n-X<\frac{1}{r}\right\} = 1 \qquad \text{para toda } n>N$$

lo anterior se denota como

$$P(\lim_{n\to\infty} X_n = X) = 1 \qquad (2)$$

o bien,

$$P\left\{\bigcup_{r=1}^{\infty} \bigcap_{N=1}^{\infty} \bigcup_{n\geq N} (w\in\Omega;\ X_n-X<\frac{1}{r})\right\} = 1.$$

Otro resultado importante es la Desigualdad de Tchebychev, esta se presenta como un caso particular del siguiente teorema, debido a que la demostración original de Tchebychev en 1867 difiere en muy pocos detalles del teorema .

Teorema 7.1.

Sea X una variable aleatoria con $E(X^2)<\infty$ entonces para cada $c>0$ se tiene que

$$P(|X|\geq c)\leq \frac{E(X^2)}{c^2}.$$

Demostración:

$$E(X^2)=\int_x x^2\, dF_X(x) = \int_{|x|\geq c} x^2\, dF_X(x) + \int_{|x|<c} x^2 dF_X(x)\geq$$

$$\geq \int_{|x|\geq c} x^2 dF_X(x) \qquad (3)$$

como $|x|\geq c$ implica que $|x|^2\geq c^2$ entonces

$$(3) \geq \int_{|x|\geq c} c^2 dF_X(x) = c^2 P(X \geq c)$$

por lo tanto

$$E(X^2)\geq c^2 P(|X|\geq c)$$

es decir

$$P(|X|\geq c)\leq E(X^2)/c^2. \qquad \blacksquare$$

Corolario 7.1. (Desigualdad de Tchebychev).

Sea $X=Y-E(Y)$ entonces resulta

$$P\left\{ |Y-E(Y)|\geq c \right\} \leq \left(\frac{1}{c^2}\, E\left\{ (Y - E(Y)^2 \right\} \right) = \frac{\sigma^2}{c^2}.$$

Por otro lado en el siglo XVIII un naturista francés Buffon lanzó una moneda 4,040 veces de las cuales obtuvo 2,048 caras, la frecuencia relativa de la ocurrencia de caras en este experimento fue de aproximadamente 0.507, Karl Pearson, lanzó una moneda 12,000 veces y obtuvo 6,019 caras, la frecuencia relativa de la ocurrencia de caras en el experimento de Pearson fue de aproximadamente 0.5016; ¿Que nos dicen estos resultados?. De acuerdo a la definición frecuentista de probabilidad, en el caso de que se tenga n eventos independientes Bernoulli, con probabilidad p de éxito, el valor de S_n/n, donde S_n es el número de éxitos en n repeticiones, se aproxima a p cuando n crece. Ahora bien, se puede dar un significado preciso de está afirmación, puesto que $E(S_n/n)= p$ y $Var(S_n/n)=Var(S_n/n^2=p(1-p)/n$ entonces de acuerdo a la Desigualdad de Tchebychev

$$P\left\{ \left|\frac{S_n}{n} - p\right| \geq c \right\} \leq \frac{1}{4nc^2}$$

Lo anterior establece que cuando n crece, la probabilidad de que el número promedio de éxitos se desvié de p en más de cualquier cantidad c asignada previamente, tiende a cero. Está afirmación es una forma de la Ley Débil de los Grandes Números. Es limitada puesto que dice que para n suficientemente grande y fija, el promedio S_n/n es muy cercano a p; pero no dice que S_n/n permanecerá necesariamente cerca de p si se incrementa el número de repeticiones. Deja abierta la posibilidad de que en n repeticiones adicionales ocurra por lo menos uno de los eventos $(S_k/k)<p-c$ con $n<k\leq 2n$. Tratando de resolver las limitaciones de la Ley Débil, surge la Ley Fuerte de los Grandes Números que establece que para $\epsilon>0$, hay probabilidad de 1 de que ocurra solamente un número finito de eventos

$$\left|\frac{S_n}{n} - p\right| \geq c \ .$$

Hasta aquí, se ha analizado las Leyes de los Grandes Numeros para un caso particular, S_n con función de distribución Bernoulli;

sin embargo, se podrá observar que se cumple esta Ley para cualesquiera variables aleatorias.

<u>Definición 7.1.</u> (Ley Débil de los Grandes Números).

Sea $(\Omega,\mathcal{A},P)$ un espacio de probabilidad, $(X_n)_{n=1}^{\infty}$ una sucesión de variables aleatorias sobre ese mismo espacio, entonces se dice que $(X_n)_{n=1}^{\infty}$ satisface *la Ley Débil de los Grandes Números* si

$$\frac{\sum_{k=1}^{n}(X_k - E(X_k))}{n} \xrightarrow{P} 0 ,$$

o bien

$$\lim_{n\to\infty} P\left\{ \left| \frac{1}{n}\sum_{k=1}^{n} X_k - \frac{1}{n}\sum_{k=1}^{n} E(X_k) \right| < c \right\} = 1.$$

<u>Teorema 7.2.</u> (Ley Débil de Tchebychev).

Si $(X_n)_{n=1}^{\infty}$ es una sucesión de variables aleatorias independientes y de varianza uniformemente acotada, entonces si para cualquier c>0

$$\lim_{n\to\infty} P\left\{ \left| \frac{1}{n}\sum_{k=1}^{n} X_k - \frac{1}{n}\sum_{k=1}^{n} E(X_k) \right| < c \right\} = 1,$$

$(X_n)_{n=1}^{\infty}$ satisface la Ley Débil de los Grandes Números.

Demostración:

Se sabe que $\mathrm{Var}\left\{\frac{1}{n}\sum_{k=1}^{n} X_k\right\} = \frac{1}{n^2}\sum_{k=1}^{n}\mathrm{Var}\,(X_k)$

y por hipótesis del teorema se tiene que la varianza es uniformemente acotada entonces $\mathrm{Var}\left\{\frac{1}{n}\sum_{k=1}^{n} X_k\right\} \le \frac{C}{n}$

aplicando la desigualdad de Tchebychev resulta que

$$P\left\{\left|\frac{1}{n}\sum X_k - \frac{1}{n}\sum E(X_k)\right| < \epsilon\right\} \geq 1 - \frac{C}{n\epsilon^2}$$,cuando $n\to\infty$ se obtiene que

$$\lim_{n\to\infty} P\left\{\left|\frac{1}{n}\sum X_k - \frac{1}{n}\sum E(X_k)\right| < \epsilon\right\} \geq 1,$$ como la probabilidad no puede exceder de uno, entonces queda demostrado el Teorema. ∎

Una consecuencia de la desigualdad de Tchebychev, es el siguiente resultado,

<u>Teorema</u> <u>7.3.</u> (Condición de Markov).

Sea (X_k) una sucesión de variables aleatorias independientes con varianza σ_k^2 $k=1,2,..$ entonces si

$$\lim_{n\to\infty} \frac{1}{n^2}\sum \sigma_k^2 = 0$$

(X_k) satisface la Ley Débil de los Grandes Números.

Demostración:

Aplicando la desigualdad de Tchebychev a la variable aleatoria $(1/n)\ \Sigma_{i=1}^{n} X_i = Y_n$ se obtiene para cada $c>0$

$$P\left\{\left| Y_n - \frac{1}{n}\sum_{k=1}^{n} E(Y_k)\right| \geq c\right\} = \frac{\Sigma_{k=1}^{n}\ \sigma_k^2}{n^2\ c^2}$$

con lo cual se establece la conclusión del teorema. ∎

En los últimos años, mucho se ha trabajado en la determinación de las condiciones que son necesarias para que las variables dependientes satisfagan las leyes de los grandes números. El teorema de Markov pertenece a esta clase de proposiciones. En este sentido, haciendo uso tanto del teorema de Tchebychev como del de Markov, se llegó a una condición que no solo es suficiente sino necesaria para que una sucesión de variables aleatorias satisfagan la ley débil de los grandes números.

<u>Teorema 7.4.</u>

Para que la sucesión $(X_n)_{n=1}^{\infty}$ de variables aleatorias no necesariamente independientes satisfaga la relación

$$\lim_{n\to\infty} P\left\{ \left| \frac{1}{n} \sum_{k=1}^{n} X_l - \frac{1}{n} \sum_{k=1}^{n} E(X_l) \right| < \varepsilon \right\} = 1$$

para cualquier número ε es necesario y suficiente que

$$E\left\{ \frac{\left[\sum_{k=1}^{n} \left(X_k - E(X_k) \right) \right]^2}{n^2 + \left[\sum_{k=1}^{n} (X_k - E(X_k)) \right]^2} \right\} \to 0$$

Demostración:

$\Leftarrow$) Sea $F_{H_n}(x) = P[H_n \leq x]$ la función de distribución de la

variable $H_n = \frac{1}{n} \sum_{k=1}^{n} (X_k - E(X_k))$

entonces

$$P\left\{ \left| \frac{1}{n} \Sigma (X_k - E(X_k)) \right| \geq \varepsilon \right\} = P\left\{ \left| H_n \right| \geq \varepsilon \right\} = \int_{|x| \geq \varepsilon} dFH_n(x)$$

por lo que

$$\leq \frac{1 + \varepsilon^2}{\varepsilon^2} \int_{|x| \geq \varepsilon} \frac{x^2}{1 + x^2} dF_{H_n}(x) \leq \frac{1 + \varepsilon^2}{\varepsilon^2} \int \frac{x^2}{1 + x^2} dF_{H_n}(x)$$

$$= \frac{1 + \varepsilon^2}{\varepsilon^2} \; E\left\{ \frac{Hn^2}{1 + Hn^2} \right\}$$

pero

$$E\left(\frac{Hn^2}{1 + Hn^2} \right) = E \left\{ \frac{\left[\sum_{k=1}^{n} \left(X_k - E(X_k) \right)^2 \right]}{n^2 + \left[\sum_{k=1}^{n} (X_k - E(X_k)) \right]^2} \right\}$$

por hipótesis

$$E \left\{ \frac{\left[\sum_{k=1}^{n} \left(X_k - E(X_k) \right)^2 \right]}{n^2 + \left[\sum_{k=1}^{n} (X_k - E(X_k)) \right]^2} \right\} \to 0$$

por lo tanto

$$P\left\{ \left| \frac{1}{n} \sum_{k=1}^{n} (X_k - E(X_k)) \right| \geq \varepsilon \right\} \to 0 \text{ cuando } n\to\infty.$$

De regreso se tiene que demostrar que $E\left\{ \frac{Hn^2}{1 + Hn^2} \right\} \to 0$ cuando $n\to\infty$.

$$\begin{aligned} P\left\{ |H_n| \geq \varepsilon \right\} &= \int_{|x|\geq\varepsilon} dF_{Hn}(x) \geq \int_{|x|\geq\varepsilon} \frac{x^2}{1 + x^2}\, dF_{Hn}(x) \\ &= \int \frac{x^2}{1 + x^2}\, dF_{Hn}(x) - \int_{|x|<\varepsilon} \frac{x^2}{1 + x^2}\, dF_{Hn}(x) \\ &\geq \int \frac{x^2}{1 + x^2}\, dF_{Hn}(x) - \varepsilon^2 = E\left\{ \frac{Hn^2}{1 + \eta n^2} \right\} - \varepsilon^2 \end{aligned}$$

por lo tanto

$$0 \leq E\left\{ \frac{Hn^2}{1 + Hn^2} \right\} - \varepsilon^2 \leq P(|H_n| \geq \varepsilon)$$

$$0 \le E\left\{ \frac{H_n^{\;2}}{1 + H_n^{\;2}} \right\} \le P(|H_n| \ge c) + c^{\;2}$$

sacando límite en ambos lados resulta

$$0 \le \lim_{n\to\infty} E\left\{ \frac{H_n^{\;2}}{1 + H_n^{\;2}} \right\} \le c^{\;2}$$

por lo tanto

$$E\left\{ \frac{H_n^{\;2}}{1 + H_n^{\;2}} \right\} \to 0 \text{ cuando } n\to\infty. \qquad \blacksquare$$

En 1909 un matemático francés llamado Emile Borel demostró un teorema, mucho más profundo que los que se tenían; éste es conocido como la Ley Fuerte de los Grandes Números. Para poder presentarlo, asícomo al teorema de Kolmogorov se introducen primero algunos conceptos necesarios para su desarrollo.

<u>Definición</u> <u>7.2.</u> (Ley Fuerte de los Grandes Números).

Sea (Ω, A, P) un espacio de probabilidad, $(X_n)_{n=1}^{\infty}$ una sucesión de variables aleatorias sobre ese mismo espacio, entonces se dice que $(X_n)_{n=1}^{\infty}$ satisface la *Ley Fuerte de los Grandes Números* si

$$\frac{\sum_{k=1}^{n} (X_k - E(X_k))}{n} \overset{cs}{\to} 0$$

o bien

$$P\left\{ \lim_{k\to\infty} \left(\frac{1}{n} \sum_{k=1}^{n} X_k \right) = \frac{1}{n} \sum_{k=1}^{n} E(X_k) \right\} = 1 .$$

Lema 7.1.

Si para todo entero positivo r $\sum_{n=1}^{\infty} P\left\{ X_n - X > \frac{1}{r} \right\}$ converge $\Rightarrow$ $X_n \to X$ (casi seguramente).

$\left(\text{esto quiere decir que } P\left\{X_n-X > \frac{1}{r}\right\} \to 0 \Leftrightarrow P\left\{X_n-X < \frac{1}{r}\right\} \to 1\right).$

Demostración:

Sea $E_n^r=(w\in\Omega;\ |X_n(w)-X(w)|\ge\frac{1}{r})$, y

$$S_n^r=\bigcup_{k=1}^{\infty} E_{n+k}^r = \bigcup_{k=1}^{\infty}(w\in\Omega: |X_{n+k}(w)-X(w)|\ge\frac{1}{r})$$

$$P(S_n^r)=P(\bigcup_{k=1}^{\infty} E_{n+k}^r) \le \sum_{k=1} P(E_{n+k}^r) = \sum_{k=1} P(w\in\Omega: |X_{n+k}(w)-X(w)|\ge\frac{1}{r})$$

por hipótesis el último término converge por lo que

$$\lim_{n\to\infty} P(S_n^r)=0 \tag{4}$$

Ahora sea $S^r= S_1^r S_2^r \ldots$, entonces

$$S^r\subset S_1^r, \ldots, S^r\subset S_n^r, \ldots,$$

por (4) resulta que

$$P(S^r)=0. \tag{5}$$

Finalmente sea $S=\bigcup_{l=1}^{\infty} S^l \Rightarrow P(S)=P(\bigcup_{l=1}^{\infty} S^l)\le\sum_{l=1}^{\infty} P(S^l)$ pero por (5) $P(S)=0$ por lo tanto

$$\begin{aligned} P(S) &= P(\bigcup_{l=1}^{\infty} S^l) = 0 \\ &\Rightarrow P(\bigcup_{l=1}^{\infty}\bigcap_{k=1}^{\infty} S_k^l) = 0 \\ &\Rightarrow P(\bigcup_{l=1}^{\infty}\bigcap_{k=1}^{\infty}\bigcup_{l=1}^{\infty} E_{k+l}^l) = 0 \\ &\Rightarrow P\left\{\bigcup_{l=1}^{\infty}\bigcap_{k=1}^{\infty}\bigcup_{l=1}^{\infty}(w\in\Omega;\ X_{k+l}(w)-X(w) \ge \frac{1}{l})\right\} = 0 \end{aligned}$$

lo cual implica que existe l tal que para toda k existe l con $l\ge k$ tal que X_{n+l} casi siempre convergue a X. ∎

Teorema 7.5. (Borel).

Sea μ el número de veces que el evento A ocurre en n intentos independientes, cada uno de los cuales ocurre con probabilidad p. Entonces para $n\to\infty$,

$$P\left\{\frac{\mu}{n}\to p\right\}=1.$$

Demostración:

De acuerdo al lema 7.1, para demostrar el teorema basta con que la serie

$$\sum_{n=1}^{\infty}P\left\{\left|\frac{\mu}{n}-p\right|\geq\frac{1}{r}\right\}$$

converga a cualquier número natural r. Tomando en cuenta lo anterior, y considerando que $E(X-E(X))^4$ existe, se define la siguiente desigualdad

$$P(|X-E(X)|\geq c)\leq\frac{1}{c^4}E(X-E(X))^4$$

entonces

$$P\left\{\left|\frac{\mu}{n}-p\right|\geq\frac{1}{r}\right\}\leq r^4E\left(\frac{\mu}{n}-p\right)^4.$$

Sea μ_i = número de ocurrencias del evento A en el i-ésimo intento. Como

$$\frac{\mu}{n}-p=\frac{1}{n}\sum_{i=1}^{n}(\mu_i-p)$$

entonces

$$E\left(\frac{\mu}{n}-p\right)^4=\frac{1}{n^4}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n}E(\mu_i-p)(\mu_j-p)(\mu_k-p)(\mu_l-p)$$

además $E(\mu_i)=p$, $\Rightarrow$ $j=i$, $k=l$, $E(\mu_j-p)^2(\mu_k-p)(\mu_l-p)$; si aparece un exponente non, el término de la suma se anula, solo cuentan en la suma cuando $i=j=k=l$, y hay n-términos, o bien, cuando $i=j$ y $k=l$, $i=l$ y $k=j$, $i=k$ y $j=l$ y hay $n(n-1)$ términos.

i) Si $\imath=j=k=\imath$ entonces

$$E(\mu_{\imath}-p)^4 = pq(p^3+q^3),$$

ii) Si $\imath=k$ y $j=\imath$ entonces

$$E\Big((\mu_{\imath}-p)^2(\mu_{j}-p)^2\Big) = p^2\,q^2,$$

luego entonces

$$E\left\{\left[\frac{1}{n}\sum(\mu_j-p)\right]^4\right\} = \frac{1}{n^4}\left[npq\,(p^3+q^3)+3n(n-1)p^2q^2\right]=$$

$$=\frac{1}{n^4}\left[npq\,(p^3+q^3+3(n-1)pq\right] \qquad (6)$$

como $pq \le \frac{1}{4} \Rightarrow (6) \le \frac{1}{4n^4}\left[npq\,(p^3+q^3+3(n-1)pq\right] \qquad (7)$

y $(p^3+q^3) \le 1$ entonces

$$(7)\le \frac{1}{4n^4}\left[n+\frac{3n^2}{4}-\frac{3n}{4}\right] = \frac{1}{4n^4}\left[\frac{3n^2}{4}+\frac{1}{4}n\right] \le$$

$$\le\frac{1}{4n^4}\left[\frac{3n^2}{4}+\frac{1}{4}n^2\right] = \frac{1}{4n^4}n^2 = \frac{1}{4n^2}$$

por lo que

$$\sum^{\infty} E\left\{\left[\frac{1}{n}\Sigma\,(\mu_j-p)\right]\right\} \le \Sigma\,\frac{1}{4n^2} < \infty$$

por lo tanto

$$\sum^{\infty} P\left\{\left[\frac{1}{n}\Sigma\,(\mu_j-p)\right]\right\} \le \Sigma\,\frac{1}{4n^2} < \infty$$

Y considerando el lema 7.1 queda demostrado el Teorema de Borel.

El Teorema anterior ha sido el punto de partida para una serie importante de investigaciones dedicadas a encontrar las condiciones bajo las cuales se cumple la Ley Fuerte de los Grandes Números. Para poder enunciar el teorema de Kolmogorov, primero se analizan la desigualdad de Kolmogorov y la Hájek-Renyi.

<u>Desigualdad de Kolmogorov.</u>

Sea X_1, $X_2,\ldots,X_n$ variables aleatorias independientes con $E(X_k^2)<\infty$, k=1,2,...,n, y sea $\eta_k = X_k - E(X_k)$; entonces para c>0,

$$P\left\{ \max_{1\leq j\leq n} |\eta_1+\ldots+\eta_j| \geq c \right\} \leq \frac{1}{\epsilon^2} \left(\sum_{j=1}^{n} \text{Var } \eta_j \right)$$

Demostración:

Considérese la cantidad

$$\zeta = S_n^2$$

donde $S_k = \sum_{j=1}^{k} \eta_j$, y cada $\eta_j = X_j - E(X_j)$. Desarrollando $E(\zeta)$ se obtiene que $E(\zeta) = \sum_{k=1}^{n} \text{Var } \eta_k$.

Sea el evento E_i (i=1,2,...,n) que consiste en el hecho de que

$$|S_j|<c \quad \text{para} \quad 1\leq j<i$$

y

$$|S_i|\geq c$$

como el evento $\max_{1\leq k\leq n} |S_k|\geq c$ implica que $|S_k|\geq c$ para algún valor de k y es equivalente al evento $\sum_{k=1}^{n} E_k$. Por ser mutuamente exclusivos resulta la siguiente igualdad

$$P(\max_{1\leq k\leq n} |S_k| \geq c) = \sum_{k=1}^{n} P(E_k).$$

Sea E_0 el evento que $S_j<c$ para $1\le j\le n$, si se considera la definición de esperanza condicional resulta que

$$E(\zeta) = \sum_{i=1}^{n} E(\zeta|E_i)\,P(E_i) \;+\; E(\zeta|E_0)\,P(E_0) \ge$$

$$\ge \sum_{i=1}^{n} E(\zeta|E_i)\,P(E_i)$$

por definición se tiene que

$$E(\zeta|E_i)=E(S_n^2|E_i)$$

pero no se conoce el valor de S_n^2, entonces si se desarrolla $(\Sigma_{j=1}\eta_j)^2$ resulta

$$E(S_n^2|E_i) = E(S_i^{\,2} + 2\sum_{j>i} S_i\eta_j + \sum_{j>i}\eta_j^{\,2} + 2\sum_{j>h>i}\eta_j\eta_h\,|E_i)$$

y

$$E(S_n^2|E_i) \ge E(S_i^{\,2} + 2\sum_{j>i} S_i\eta_j + 2\sum_{j>h>i}\eta_j\eta_h\,|E_i)$$

analizando cada término con respecto a la ocurrencia de E_i se observa que solo impone restricciones para las primeros i de las variables aleatorias η_j las subsecuentes permanecen independientes unas de otras y de S_i. Por lo anterior se tiene que para

$$j>i \qquad E(S_i\eta_j|E_i) = E(\eta_j|E_i)\,E(S_i|E_i) = 0$$
$$j>h>i \qquad E(\eta_j\eta_h|E_i) = 0$$

entonces resulta que

$$E(S_n^2|E_i) \ge E(S_n^2|E_i)$$

como se realiza el evento E_i entonces

$$E(S_n^2|E_i) \geq c^2$$

por lo tanto

$$E(\zeta|E_i) \geq c^2$$

y

$$E(\zeta) \geq c^2 \sum_{i=1}^{n} P(E_i)$$

con lo cual queda demostrada la desigualdad. ∎

Desigualdad de Hájek-Renyi.

La demostración de la desigualdad se basa en la demostración de la desigualdad de Kolmogorov . Por lo que solo se enuncia:

Sea $X_1, X_2, \ldots, X_n$ variables aleatorias independientes con $E(X_k^2)<\infty$, $k=1,2,\ldots,n$, y sea $\eta_k = X_k - E(X_k)$ con $E(\eta_k)=0$; si $c_k (k=1,2,\ldots)$ es una sucesión no decreciente de constantes positivas entonces para cualesquiera $m<n$ y $c>0$,

$$P\left\{ \max_{m\leq j\leq n} |\eta_1+\ldots+\eta_j| \geq c \right\} \leq \frac{1}{\epsilon^2}\left(c_m^2 \sum_{j=1}^{m} \mathrm{Var}\,\eta_j + \sum_{j=m+1}^{n} c_k^2 \ \mathrm{Var}\ \eta_j \right).$$

Teorema 7.6. (Kolmogorov).

Si una sucesión de variables aleatorias independientes satisface

$$\sum_{n=1}^{\infty} \frac{\mathrm{Var}\ (X_n)}{n^2} < \infty$$

entonces la sucesión $(X_n)_{n=1}^{\infty}$ cumple con la Ley Fuerte de los Grandes Números.

Demostración:

Considérese la desigualdad de Hájek-Rényi, sustituyendo $c_k=1/k$ entonces

$$P\left\{ \max_{m\le j\le n} \frac{1}{k} \left| \sum_{s=1}^{k} \eta_s \right| \ge c \right\} \le \frac{1}{\epsilon^2} \left(\frac{1}{m^2} \sum_{j=1}^{m} \mathrm{Var}\ \eta_j + \sum_{j=m+1}^{n} \frac{\mathrm{Var}\ \eta_j}{j^2} \right)$$

si $n \to \infty$ entonces queda

$$P\left\{ \max_{m\le j} \frac{1}{k} \left| \sum_{s=1}^{k} \eta_s \right| \ge c \right\} \le \frac{1}{\epsilon^2} \left(\frac{1}{m^2} \sum_{j=1}^{m} \mathrm{Var}\ \eta_j + \sum_{j=m+1}^{\infty} \frac{\mathrm{Var}\ \eta_j}{j^2} \right)$$

si se considera que $\Sigma_{j=1}^{\infty}$ Var η_j/k^2 converge entonces de esta última desigualdad se desprende que

$$\lim_{m\to\infty} P\left\{ \max_{m\le j} \frac{1}{k} \left| \sum_{s=1}^{k} \eta_s \right| \ge c \right\} = 0$$

lo cual significa que

$$P\left\{ \lim_{m\to\infty} \frac{1}{k} \left| \sum_{s=1}^{k} \eta_s \right| = 0 \right\} = 1$$

y esto implica que se cumple con la Ley Fuerte de los Grandes Números. ∎

Teorema del Límite Central.

Ahora, el problema que se presenta es el de analizar las leyes que caracterizan a la suma de un número grande de variables aleatorias cada una de las cuales tiene un pequeño efecto en la suma. Considérese los siguientes ejemplos; supóngase que se realizan n visitas a un consultorio médico, sea X que represente el número de veces que se encuentre al doctor, entonces ¿Cuál es la probabilidad de que en más de la mitad de las veces, se encuentre al doctor?. En este caso se tendría que calcular la $P[S_n>n/2]$ donde $S_n=\sum_{i=1}^{n} X_i$. O bien sea X que represente la vida de una pila, utilizada en un radio, con distribución exponencial y media de 20 días. Tan pronto como se acaba la pila, se repone con una nueva. ¿Cuál es probabilidad de que se necesiten más de 25 pilas en un año?. En este caso se tendría que calcular $P[S_{25}<365]$ donde $S_n=\sum_{i=1}^{n} X_i$. ¿Cómo se determinaría la función de distribución de la suma?. Antes de enunciar el Teorema del Límite Central se introduce el concepto de convergencia de funciones de distribución.

Definición 7.3.

Se dice que una sucesión de funciones de distribución $(F_n(x))$ n=1,2,... converge a una función de distribución límite $F(x)$, si en cualquier punto de continuidad de $F(x)$, se tiene que el $\lim_{n\to\infty} F_n(x)=F(x)$. Por simplificación, se denota a este tipo de convergencia como $F_n(x) \xrightarrow{d} F(x)$.

El hecho de pedir que converga en todos los puntos de continuidad de la función límite,es un requisito muy fuerte para probar la convergencia en distribución, por lo que el siguiente teorema trata de eliminar esta condición.

Teorema 7.7.

Sea $F_n(x)$ una sucesión de funciones de distribución y sea $\phi_n(t)$ su correspondiente función característica. Entonces $F_n(x)\to F(x) \Leftrightarrow \lim_{n\to\infty} \phi(t) = g(t)$ para toda t, donde $g(t)$ es continua en t=0. Además, $g(t)=\phi(t)$, la función característica de $F(x)$.

La demostración se omite; pero para mayores detalles se debe referir a E. Luckacs, "Characteristics Functions".

Ahora se analiza el *Teorema del Límite Central*; este resultado permite considerar a la función de distribución límite como la solución al problema en lugar de analizar las funciones de distribución de las sumas de las variables aleatorias con un gran número de términos.

Sea (X_k) $(k=1,2,\ldots)$ una sucesión de variables aleatorias, y si existe una sucesión de constantes α_n y β_n $(n=1,2\ldots)$ tales que para

$$U_n = \sum_{k=1}^{n} X_k - \alpha_n / \beta_n, \qquad n=1,2,\ldots$$

se tiene que

$$F_{U_n}(x) \xrightarrow{d} \int_{-\infty}^{x} \frac{1}{\sqrt{2\Pi}} e^{-t^2/2} dt$$

entonces (X_k) cumple el Teorema del Límite Central.

Una cualidad importante del Teorema del Límite Central es el hecho de que las variables aleatorias (X_k) se pueden aproximar con una función de distribución normal sin importar cual sea su función de distribución original.

Como la función de distribución límite es una normal con media 0 y varianza 1, y si para cada k, $E(X_k)=\mu_k$ y $E(X_k-\mu_k)^2=\sigma_k^2$ existe y las variables aleatorias (X_k) son mutuamente independientes, entonces sería natural escoger a α_n y β_n como:

$$\alpha_n = \sum_{k=1}^{n} \mu_k \qquad \text{y} \qquad \beta_n = \sqrt{\Sigma_{k=1}^{n} \sigma_k^2}$$

entonces

$$U_n = \sum_{k=1}^{n} X_k - \mu_k / \sqrt{\Sigma_{k=1}^{n} \sigma_k^2}$$

y si se cumple

$$F_{U_n}(x) \xrightarrow{d} \int_{-\infty}^{x} \frac{1}{\sqrt{2\Pi}} e^{-t^2/2} dt$$

entonces se dice que (X_k) cumple con el Teorema del Límite Central Clásico.

<u>Teorema 7.8.</u>

Sea $X_1, X_2, \ldots, X_n$ variables aleatorias con función de distribución $F(x)$ y sea $E(X)=\mu$, y $\sigma^2_x=\sigma^2$. Entonces (X_k) cumplen con el Teorema del Límite Central. Donde $U_n=\sqrt{n}\,(\bar{X}-\mu)/\sigma$.

Demostración:

Se reescribe $U_n=\dfrac{X_n - n\mu}{\sigma\sqrt{n}}$ en donde $X_n=\Sigma X_i$ entonces para t fija y n suficientemente grande,

$$\varphi_{U_n}(t)=e^{-in\mu t/\sigma\sqrt{n}}\,\varphi_{X_n}(t/\sigma\sqrt{n})$$
$$=e^{-in\mu t/\sigma\sqrt{n}}\left(\varphi_X(t/\sigma\sqrt{n})\right)^n$$

ó

$$\varphi_{U_n}(t)=\exp\left(n(\log\varphi_X(t/\sigma\sqrt{n}) - i\mu(t/\sigma\sqrt{n}))\right)$$

ahora se calcula el límite y se obtiene que

$$\lim_{n\to\infty} n(\log\varphi_X(t/\sigma\sqrt{n}) - i\mu(t/\sigma\sqrt{n})) = -\frac{t^2}{2}$$

si t=0 se satisfacen ambos lados de la igualdad; si t≠0 entonces se reescribe el lado derecho del límite y resulta que,

$$\frac{t^2}{\sigma^2}\lim_{n\to\infty}\frac{\log\varphi_x(t/\sigma\sqrt{n}\,) - i\mu(t/\sigma\sqrt{n}\,)}{(t/\sigma\sqrt{n}\,)^2}$$

por otro lado se tiene que si se aplica la ley de Hôpital se tiene que

$$\lim_{t\to 0}\frac{\log\varphi_x(t) - i\mu t}{t^2} = -\frac{\sigma^2}{2}$$

entonces si $t\neq 0$ el límite resulta en

$$\frac{t^2}{\sigma^2}\left(-\frac{\sigma^2}{2}\right) = -\frac{t^2}{2}$$

entonces

$$\lim_{n\to\infty}\varphi_{U_n}(t) = e^{-t^2/2} \qquad -\infty<t<\infty$$

que es la función característica de una variable aleatoria con función de distribución normal. ∎

Ahora se enuncia el Teorema de Lindeberg-Feller, el cual impone condiciones necesarias y suficentes para que una sucesión de variables aleatorias mutuamente independientes satisfaga el Teorema del Límite Central.

<u>Teorema 7.9.</u> (Lindeberg-Feller).

Sea $\{X_k\}$ k=1,2,... una sucesión de variables aleatorias mutuamente independientes con media μ_k y varianza σ_k^2 respectivamente. Entonces una condición necesaria y suficiente para que la sucesión de variables aleatorias $\{X_k\}$ cumplan con el Teorema del Límite Central y para que

$$\lim_{n\to\infty} \max_{j} \frac{\sigma_j^2}{\Sigma_{j=1}^n \sigma_j^2} = 0$$

es que para cualquier $\varepsilon>0$,

$$\lim_{n\to\infty} (1/\Sigma_{j=1}^n \sigma_j^2) \sum_{k=1}^n \int_{|x-\mu_k|\le\varepsilon\sqrt{\Sigma_{j=1}^n \sigma_j^2}} (x-\mu_k)^2 dF_k(x) = 1.$$

La condición de suficiencia fué establecida en 1922 por J.W. Lindeberg, y la necesidad fué probada por W. Feller en 1935.

La demostración es muy compleja para incluirla en el presente trabajo.[1]

Teorema 7.10. (Lyapunov).

Sea (X_k) $(k=1,2,\ldots)$ una sucesión de variables aleatorias mutuamente independientes tales que para algún $\delta>0$, $E(|X_k-\mu_k|^{2+\delta})$ existe para cada k. Entonces si

$$\lim_{n\to\infty} (1/\Sigma_{j=1}^n \sigma_j^2)^{1+\delta/2} \sum_{k=1}^n E\left\{ |X_k-\mu_k|^{2+\delta} \right\} = 0$$

(X_k) cumple con el Teorema del Límite Central.

Demostración:

Para cualquier $\varepsilon>0$

$$(1/\Sigma_{j=1}^n \sigma_j^2) \sum_{k=1}^n \int_{|x-\mu_k|>\varepsilon\sqrt{\Sigma_{j=1}^n \sigma_j^2}} (x-\mu_k)^2 dF_k(x)$$

---

[1] Referirse a Probability Theory-M. Loéve.

$$\leq \left(1/\left(\sum_{j=1}^{n}\sigma_j^2\right)\left(\varepsilon\sqrt{\sum_{j=1}^{n}\sigma_j^2}\right)^{\delta}\right)$$

$$\sum_{k=1}^{n}\int_{|x-\mu_k|>\varepsilon\sqrt{\sum_{j=1}^{n}\sigma_j^2}} x-\mu_k^{2+\delta} \quad dF_k(x)$$

entonces

$$(1/\sum_{j=1}^{n}\sigma_j^2)\sum_{k=1}^{n}\int_{|x-\mu_k|>\varepsilon\sqrt{\sum_{j=1}^{n}\sigma_j^2}} (x-\mu_k)^2 dF_k(x)$$

$$\leq (1/\varepsilon^{\delta}(\sum_{j=1}^{n}\sigma_j^2)^{1+\delta/2}\sum_{k=1}\int_{-\infty}^{\infty}|X_k-\mu_k|^{2+\delta}\,dF_k(x)$$

el lado derecho puede escribirse como

$$(1/\varepsilon^{\delta}\sum_{j=1}^{n}\sigma_j^2)^{1+\delta/2}\sum_{k=1}E\left\{\ |X_k-\mu_k| \quad ^{2+\delta}\right\}$$

que por hipótesis tiende a cero cuando $n\to\infty$. Por lo tanto el lado derecho se aproxima a cero y

$$(1/\sum_{j=1}^{n}\sigma_j^2)\sum_{k=1}^{n}\int (x-\mu_k)^2 dF_k(x) = 1 . \qquad \blacksquare$$

Ejemplo:

Veinte números son redondeados al entero más cercano y luego se suman. Considérese que el error de cada redondeo es independiente y uniformemente distribuido sobre el intervalo (-1/2,1/2) . ¿Cuál es la probabilidad de que la suma dada difiera de la suma original de los veinte números en más de tres?.

Sea X_i el error i-ésimo de cada $X_i\sim U(-1/2,1/2)$ entonces lo que se desea encontrar es

$P(X_1+X_2+...+X_{20} \geq 3)=1-P(X_1+...+X_{20}) \leq 3)$

$$=1-P(-3 \leq X_1+...+X_{20} \leq 3)$$

$$=1-\left\{P((X_1+...+X_{20}) \leq 3) - P((X_1+...+X_{20}) \leq -3)\right\}$$

como cada X_i se distribuye uniformemente implica que

$$f_x(x) = \begin{cases} 1 & \text{si } x \in (-1/2, 1/2) \\ 0 & \text{en otro caso} \end{cases}$$

por lo tanto

$$E[X]=\int_{-1/2}^{1/2} x\, dx = \frac{x^2}{2}\Bigg|_{-1/2}^{1/2} = 0$$

y

$$\mathrm{Var}[X]=E[X^2]=\int_{-1/2}^{1/2} x^2 dx = \frac{x^3}{3}\Bigg|_{-1/2}^{1/2} = \frac{1}{24} - \left(-\frac{1}{24}\right) = \frac{1}{12},$$

si se aplica el teorema del límite central resulta que

$$1-\left\{P((X_1+...+X_{20}) \leq 3) - P((X_1+...+X_{20}) \leq -3)\right\} =$$

$$= 1 - \left(\left(\Phi\frac{3}{\sqrt{5/3}}\right) - \Phi\frac{-3}{\sqrt{5/3}}\right)$$

buscando los resultados en las tablas de la función de distribución normal resulta que

$$= 1 - (.9898 - 0.012)$$
$$= 0.0204$$

Por lo tanto la probabilidad de que la suma dada difiera de la original en más de tres es 0.0204.

A fin de dar una introducción a la Teoría de Probabilidad, se compilaron sus conceptos más generales, buscando en todo momento, presentarlos en una forma intuitiva. Este hecho, aunado a que se recoge la experiencia docente de especialistas en el ramo, el documento constituye una herramienta ágil para utilizarse como apoyo didáctico en el que se establece una base de conocimientos sin una rígida teoría.

Para lograrlo, se expone inicialmente el marco de estudio de la probabilidad, se presentan algunas de sus definiciones poniendo especial interés en la Definición Axiomática, puesto que con ella se establece un marco matemático formal para su estudio. Así, para determinar la probabilidad dentro de este marco, en el segundo capítulo se comentaron los conceptos de variable aleatoria y función de distribución; se presentó también la generalización de estas ideas a varias variables aleatorias ayudados de los conceptos de vector aleatorio y función de distribución conjunta en el tercer capítulo.

Tomando en consideración el hecho de que en algunas ocasiones se presenta la necesidad de encontrar la función de distribución de una variable aleatoria que se encuentra expresada en término de otras variables aleatorias en las que se conoce su función de distribución, en el cuarto capítulo se incluyó un estudio de la transformación de variables aleatorias.

Ahora bien, no obstante que la función de distribución es una característica completa de la variable aleatoria, a veces es de interes conocer tan solo una característica numérica más simple, como son su esperanza, varianza, covarianza, o su coeficiente de correlación; así en el quinto capítulo se tratarón a las características numéricas tanto de variables como vectores aleatorios.

Dentro de la teoría probabilidad se reconoce a la función característica como una herramienta muy útil, conociendola se puede determinar tanto la función de distribución de la variable aleatoria, como sus características numéricas. Además considero que es un método menos laborioso que el de las transformaciones en la obtención de la función de distribución de variables aleatorias definidas en término de otras variables en las que se conoce su distribución, en el sentido de que basta con identificar la función característica resultante para relacionarla con su correspondiente función de distribución. Es por esto, que en el sexto capítulo se presentó un estudio sobre la función característica y en el apéndice se incluyen algunas funciones características.

En el último capítulo se tratan dos temas muy importantes la Ley de los Grandes Números y el Teorema del Límite Central. Ambos temas estudian las regularidades que se presentan cuando la repetición de los eventos crece indefinidamente. El primero se puede entender como el conjunto de teoremas que establecen que si un evento tiene probabilidad positiva sin importar que tan pequeña sea está, y si la probabilidad de ocurrencia es la misma en cada intento y si el número de intentos es muy grande, entonces la probabilidad de que ocurra al menos una vez, puede ser tan cercano a uno como se desee. Con el segundo se estudian las leyes que caracterizan a la suma de un número grande de variables aleatorias. Una cualidad importante del Teorema del Límite Central es el hecho de que establece la forma para que $\{X_k\}_{k=1,...,n}$ una sucesión de variables aleatorias se puedan aproximar con una función de distribución normal sin importar cual sea su función de distribución original.

Los temas presentados constituyen una base de conocimientos, que permite estudiar y entender através de la probabilidad el comportamiento de diversos fenómenos que se presentan en diversas disciplinas del conocimiento humano. Con el deseo de que este esfuerzo sea de utilidad a futuras generaciones.

APENDICE

I. VARIABLES ALEATORIAS DISCRETAS.

Uniforme (n).

a) Función de densidad,

$$f_x(x) = \begin{cases} \frac{1}{n} & \text{si } x = 1,2,\ldots n \\ 0 & \text{en otro caso} \end{cases}$$

b) Función Generatriz de Momentos,

$$m(t) = \sum_{j=1}^{n} e^{tj} \frac{1}{n}$$

c) Función Característica,

$$\varphi_x(t) = \sum_{j=1}^{n} e^{itj} \frac{1}{n}.$$

Bernoulli (p).

a) Función de densidad,

$$f_X(x) = \begin{cases} p^x(1-p)^{1-x} & x=0,1 \\ 0 & \text{en otro caso} \end{cases}$$

b) Función Generatriz de Momentos,

$$m(t) = (1-p) + pe^t$$

c) Función Característica,

$$\varphi_x(t) = (1-p) + pe^{it}.$$

Binomial (n,p).

a) Función de densidad,

$$f_x(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & x=0,1,\ldots,n \\ 0 & \text{en otro caso} \end{cases}$$

b) Función Generatriz de Momentos,

$$m_x(t)=(pe^t+(1-p))^n$$

c) Funcion Característica,

$$\varphi_x(t)=(pe^{it}+(1-p))^n.$$

Poisson (λ).

a) Función de densidad,

$$f_X(x) = \begin{cases} \dfrac{\lambda^x e^{-\lambda}}{x)} & x=0,1,2,\ldots \\ 0 & \text{en otro caso} \end{cases}$$

b) Función Generatriz de Momentos,

$$m_x(t)=\exp\left(\lambda(e^t-1)\right)$$

c) Función Característica,

$$\varphi_x(t)=\exp\left(\lambda(e^{it}-1)\right).$$

Binomial Negativa (r,p).

a) Función de densidad

$$f_x(x)=\begin{cases}\binom{r+x-1}{x}p^r(1-p)^x & x=0,1,2,\ldots\\ 0 & \text{en otro caso}\end{cases}$$

b) Función Generatriz de Momentos,

$$m_x(t)=\left(\frac{p}{1-(1-p)e^t}\right)^r \qquad \text{si } e^t(1-p)<1$$

c) Función Característica,

$$\varphi_x(t)=\left(\frac{p}{1-(1-p)e^{it}}\right)^r \qquad \text{si } e^{it}(1-p)<1.$$

Geométrica (p).

a) Función de densidad,

$$f_x(x)=\begin{cases}p(1-p)^x & x=0,1,2,\ldots\\ 0 & \text{en otro caso}\end{cases}$$

b) Función Generatriz de Momentos,

$$m_x(t)=\frac{p}{1-(1-p)e^t} \qquad \text{si } e^t(1-p)<1$$

c) Función Característica,

$$\varphi_x(t)=\frac{p}{1-(1-p)e^{it}} \qquad \text{si } e^{it}(1-p)<1.$$

II. VARIABLES ALEATORIAS CONTINUAS.

Uniforme (a,b).

a) Función de densidad,

$$f_X(x) = \begin{cases} \dfrac{1}{b-a} & a<x<b \\ 0 & \text{en otro caso} \end{cases}$$

b) Función Generatriz de Momentos,

$$m_x(t)=\frac{e^{bt}-e^{at}}{(b-a)t}$$

c) Función Característica,

$$\varphi_x(t)=\frac{e^{ibt}-e^{iat}}{(b-a)it}.$$

Normal (μ,σ^2).

a) Función de densidad,

$$f_X(x) = \frac{1}{\sqrt{2\pi}\,\sigma}\exp\left[\frac{-(X-\mu)^2}{2\sigma^2}\right] \quad \text{con } x\in\mathbb{R}$$

b) Función Generatriz de Momentos,

$$m_X(t)=\exp\left(t\mu+\frac{t^2\sigma^2}{2}\right)$$

c) Función Característica,

$$\varphi_X(t)=\exp\left(it\mu - \frac{t^2\sigma^2}{2}\right).$$

Gamma (α,λ).

a) Función de densidad,

$$\Gamma(\alpha,\lambda)=\begin{cases}\dfrac{\lambda^{\alpha}x^{\alpha-1}e^{-\lambda x}}{\Gamma(\alpha)} & x>0,\ \lambda>0,\ \alpha>0\\ 0 & \text{en otro caso}\end{cases}$$

b) Función Generatriz de Momentos,

$$m_x(t)=\left(\frac{\lambda}{\lambda - t}\right)^{\alpha} \qquad \text{para } t<\lambda$$

c) Función Característica,

$$\varphi_x(t)=\left(\frac{\lambda}{\lambda - it}\right)^{\alpha} \qquad \text{para } it<\lambda .$$

Exponencial (λ).

a) Función de densidad,

$$f(x) = \begin{cases}\lambda e^{-\lambda x} & x>0,\ \lambda>0,\\ 0 & \text{en otro caso}\end{cases}$$

b) Función Generatriz de Momentos,

$$m_x(t)=\left(1-\frac{t}{\lambda}\right)^{-1}$$

c) Función Caracteristica,

$$\varphi_x(t)=\left(1-\frac{t}{\lambda}\right)^{-1}.$$

<u>Chi-Cuadrada (k).</u>

a) Función de densidad,

$$f_X(x)=\frac{1}{\Gamma(k/2)}\left(\frac{1}{2}\right)^{k/2}x^{k/2-1}\,e^{-1/2X}\qquad k=1,2,\ldots$$

b) Función Generatriz de Momentos,

$$m_x(t)=(1-2t)^{-k/2}$$

c) Función Característica,

$$\varphi_x(t)=(1-2it)^{-k/2}.$$

<u>Beta (n,m).</u>

a) Función de densidad,

$$B(n,m)=\begin{cases}\dfrac{\Gamma(m+n)\,x^{n-1}(1-x)^{m-1}}{\Gamma(m)\Gamma(n)} & x\in(0,1)\quad m,n>0\\ 0 & \text{en otro caso}\end{cases}$$

b) Función Generatriz de Momentos,

$$m_x(t)=\frac{\Gamma(m+n)}{\Gamma(m)\Gamma(n)}\int_0^1 e^{tx}\, x^{n-1}(1-x)^{m-1}dx$$

c) Función Característica,

$$\varphi_x(t)=\frac{\Gamma(m+n)}{\Gamma(m)\Gamma(n)}\int_0^1 e^{itx}\, x^{n-1}(1-x)^{m-1}dx\ .$$

<u>Cauchy.</u>

a) Función de densidad,

$$f_x(x)=\frac{1}{\Pi(1+x^2)} \qquad -\infty<x<\infty$$

b) No tiene función generatriz de momentos,

c) Función Característica,

$$\varphi_x(t)=e^{|t|}.$$

<u>Doble</u> <u>Exponencial</u> (σ,β).

a) Función de densidad,

$$f_x(x)=\frac{1}{2\beta}\exp\left(-\frac{x-\sigma}{\beta}\right) \qquad -\infty<\sigma<\infty,\ \beta>0$$

b) Función Generatriz de Momentos,

$$m_x(t)=\frac{e^{\sigma t}}{1-(\beta t)^2}$$

c) Función Característica,

$$\varphi_x(t)=\frac{e^{i\sigma t}}{1-(\beta i t)^2}\ .$$

BIBLIOGRAFIA

- Feller William, Editorial LIMUSA 1983.
  Introducción a la Teoría de Probabilidades y sus Aplicaciones.

- Gnedenko B. V., Chelsea Publishing Company 1962.
  The Theory of Probability.

- Harris Bernard, Addison-Wesley Publishing Company 1966.
  Theory of Probability.

- Hérault Daniel, Ediciones de Promoción Cultural, S.A. 1973.
  Elementos de Teoría de la Probabilidad.

- Hoel Port Stone, Houghton Mifflin Company 1971.
  Introduction to Probability Theory.

- M. Loéve., Third Edition, D. Van Nostrand Company 1963.
  Probabililty Theory.

- Mood Graybill Boes, Mc. Graw Hill 1974.
  Introduction to the Theory of Statistics.

- Pugachev V. S., Editorial Mir Moscú 1973.
  Introducción a la Teoría de las Probabilidades.